THE PHYSICAL SOCIETY

Reports on
PROGRESS IN PHYSICS

VOLUME II

GENERAL EDITOR
ALLAN FERGUSON

1935

Reprinted 1970 for
Wm. DAWSON & SONS LTD., LONDON
with the permission of
THE PHYSICAL SOCIETY

Originally printed in Great Britain at
The Cambridge University Press
Reprinted by Cotswold Litho Ltd., Eynsham, Oxford

CONTENTS

PAGE

General physics (classical). By ALLAN FERGUSON 1

The quantum theory. (*Edited by* Prof. G. TEMPLE) 27

PART 1. Quantum theory of atomic nuclei. By R. PEIERLS

PART 2. The spectra of polyatomic molecules. By G. B. B. M. SUTHERLAND

PART 3. The quantum theory of valency. By W. G. PENNEY

PART 4. The theory of paramagnetism and related properties in ionic crystals. By R. SCHLAPP and W. G. PENNEY

Atomic physics. By N. FEATHER 74

Geophysical prospecting. By E. LANCASTER-JONES 97

Radio exploration of upper-atmospheric ionization. By E. V. APPLETON . 129

Sound. By E. G. RICHARDSON 166

Heat. By J. H. AWBERY 188

Electrical and magnetic measurements. By L. HARTSHORN 221

The charge of the electron. By Prof. H. R. ROBINSON 247

Electron tubes. (*Edited by* Prof. G. I. FINCH) 258

PART 1. General introduction. By Prof. G. I. FINCH

PART 2. The high-speed cathode-ray oscillograph. By J. L. MILLER and J. E. L. ROBINSON

PART 3. The electron microscope. By H. J. H. STARKS

PART 4. The electron-diffraction camera. By G. I. FINCH

X rays. By THE X-RAY STAFF, PHYSICS DEPARTMENT, NATIONAL PHYSICAL LABORATORY. (*Edited by* G. W. C. KAYE) 301

§1. Production of X rays. By G. E. BELL

§2. Ionization in gases, absorption and scattering of X rays. By I. BACKHURST

§3. X-ray (and gamma-ray) quantity or dose. Protection. By W. BINKS

§4. Radium. By W. E. PERRY

§5. X-ray spectroscopy. By G. SHEARER

§6. Crystal structure of metals and inorganic substances. By W. A. WOOD

§7. Crystal structure of organic and biological substances. By J. THEWLIS

PAGE

Spectroscopy. (*Edited by* HERBERT DINGLE) 329

§ 1. The determination of wave-length standards. By C. V. JACKSON

§ 2. Isotope effects in the spectra of diatomic hydrides. By R. W. B. PEARSE

§ 3. Astronomical spectroscopy. By A. HUNTER

§ 4. Infra-red band spectra of organic molecules. By LOTTE KELLNER

§ 5. The hyperfine structure of spectrum lines and nuclear moments. By E. GWYNNE JONES

Optics. (*Edited by* T. SMITH) 356

§ 1. The theory of optical instruments. By T. SMITH

§ 2. The teaching of elementary optics. By T. SMITH

§ 3. Vision. By J. GUILD

§ 4. Colorimetry. By J. GUILD

GENERAL PHYSICS (CLASSICAL)

By ALLAN FERGUSON

§ 1. INTRODUCTION

It is quite impossible, in a survey which is strictly limited in the number of pages which it may occupy, to cover in any detail the many branches of general physics which call for comment. It seemed to the writer, therefore, preferable to give fairly full details concerning selected papers in one or two branches of this subject, choosing the papers for their general interest and importance, and paying due regard to the experimental and theoretical aspects of the matters studied. In making such a choice, personal interests necessarily play their part, but the writer has endeavoured to discuss advances which should be of general interest and should be helpful in suggesting lines for further research and criticism. In this section of the report, attention has been directed to certain advances in the subjects of viscosity, surface tension, general dynamics, and Brownian movement. In the third volume of the Reports, it is hoped to include in this section a review of certain problems of elasticity, and possibly to include a special review of recent work on adsorption.

§ 2. BROWNIAN MOTION

The early history of the Brownian motion is told in many easily accessible treatises[1]; its interest for the present generation of enquirers dates from the derivation of the mean square value of the displacement in the classic paper by Einstein[2,3], this quantity being of primary interest as a magnitude capable of direct observation. The equation which gives this mean square value ($\bar{s}^2$) is, in its simplest form,

$$\bar{s}^2 = 2kTt/f \qquad \text{.....(i)},$$

f being a frictional coefficient, and k the Boltzmann constant.

The basis of all such formulae is a simple equation of motion which assumes a retarding force proportional to the velocity and which, if we restrict ourselves to a consideration of a component motion parallel to the axis of x, will be

$$X - f\dot{x} = m\ddot{x} \qquad \text{.....(ii)},$$

where we see that the forces on the particle of mass m due to the medium are split into two factors—a steady frictional part ($f\dot{x}$) which, in a more general case may be written as some function $\phi(\dot{x})$ of the velocity, and a fluctuating part X which is, in general, a function of the time. More generally, then, the fundamental equation of motion of the particle may be written

$$F(t) - \phi(\dot{x}) = m\ddot{x}.$$

Multiplying (ii) by x and remembering that

$$x\ddot{x} = \frac{1}{2}\frac{d^2}{dt^2}(x^2) - \dot{x}^2,$$

we easily see that

$$xX - \frac{f}{2}\frac{d}{dt}(x^2) = \frac{m}{2}\frac{d^2}{dt^2}(x^2) - m\dot{x}^2,$$

which equation, summed at any one instant over the whole system of particles, gives at once

$$\frac{m}{2}\frac{dz}{dt} - \frac{RT}{N} = -f\frac{z}{2}.$$

Here, z stands for $\frac{d}{dt}(\bar{x}^2)$, $\bar{x}^2$ being the mean of the different values of x^2. The equation takes this form n virtue of the facts that ΣXx vanishes and that the mean kinetic energy of a particle may be written (remembering that we are considering the component motion along the x-axis), as one-third of the usual value $3RT/2N$.

Integrating this equation we find that, in ordinary circumstances, for values of the time of the order of seconds (and indeed much less) the exponential term which appears in the integral is quite negligible, so that

$$z = d(\bar{x}^2)/dt = 2RT/Nf.$$

Hence if $\bar{s}^2$ represents the *change* in $\bar{x}^2$ in an interval t,

$$\bar{s}^2 = 2RTt/Nf,$$

which is equation (i)[4]. The equation unfolds the history of the change in $\bar{s}^2$ for n particles in a time t. The usual and convenient way of using the equation is to fix attention on one individual particle and to observe its displacements through n time-intervals t. The form of the friction coefficient f depends on the circumstances of the individual case considered. If the particles are immersed in a liquid or in a gas at ordinary pressure, the mean free paths of the molecules of the medium are small in comparison with the mean free paths of the particles in Brownian motion. In this case the medium—as far as the steady frictional forces are concerned—may be treated as a continuum, and, as was assumed by Einstein, the Stokes equation may be applied, giving $6\pi\eta a$ as the value of the frictional coefficient.

Should the surrounding medium be a highly rarefied gas, in which the mean free paths of the molecules of the medium are *greater* than those of the particles in Brownian motion, then the Stokes friction law has to be replaced by what is termed a Doppler friction; this leads to a quite determinate solution, the calculation of the Doppler friction coefficient being based on the fact that a particle moving say, to the right, will receive more impacts from the right than from the left.

A number of papers dealing with the theory of the Brownian movement have been published during the last few years. Uhlenbeck and Ornstein[5] show that, where u is the velocity and s the displacement of a free particle in Brownian motion, the quantities

$$u - u_0 e^{-\beta t} \quad \text{and} \quad s - u_0/\beta\,(1 - e^{-\beta t})$$

(u_0 being the initial velocity and β the friction coefficient per unit mass) follow the normal Gaussian distribution law. For s this gives the exact frequency distribution corresponding to the exact formula for $\bar{s}^2$ given by Ornstein and Furth[6].

Uhlenbeck and Goudsmit[7] discuss the implications of some experimental work by Gerlach[8] who studied photographically the Brownian motion of a minute mirror attached to a fine wire. The behaviour of such a system, having but one characteristic frequency, can be analysed without great difficulty. Gerlach studied the phenomena over a wide range of pressure and found that the mean square deviation of the mirror was a function of the temperature only of the surrounding gas. The authors, taking as the equation of motion

$$\ddot{\phi} + r\dot{\phi} + \omega^2\phi = F(t),$$

where $F(t)$ is the fluctuating torque per unit moment of inertia due to the impacts of the gas molecules, develop this function in a Fourier's series over a time-interval which is great compared with the period of the mirror. The authors find that the square of the amplitude of *each* of the Fourier components is a function of the pressure, the temperature and the molecular weight of the gas surrounding the mirror, but that the sum of the squares is a function of the temperature only. By a very simple analysis they develop appropriate expressions for the friction coefficient at low pressures and obtain a simple final value for the mean square angular deviation.

Metadier[9, 10] has developed an operational method for the study of Brownian motion, and has considered the relation of his method of attack to the classical method. He concludes that the methods are equivalent, and that, according to the circumstances one or the other method may be convenient.

Sato[11] has investigated mathematically the influence of heat flow on Brownian motion. He considers the Brownian motion of a particle in a gas between two surfaces maintained at different temperatures, and shows that at very low pressures the heat flow has a considerable effect on the motion, but that the effect is negligible at ordinary pressures. The same author[12] has discussed the theory of Brownian motion from the point of view of Fermi-Dirac statistics, and shows that Brownian motion in the Einstein sense takes place only when the gas is not in a quantized state. For a gas in a strongly quantized state a new result, essentially different from the classical one, is deduced, and it is shown that Planck's constant may be deduced from the observation of Brownian motion in these circumstances.

Ornstein and van Wijk[13], by integration of Langevin's equation, have derived the distribution functions of velocity and displacement for a free and for a harmonically bound particle. The results are extended to include the influence of an external force dependent on the time.

A number of interesting pieces of experimental work fall to be chronicled. Brumberg and Vavilov[14, 15] describe a new method to test the laws of Brownian motion. The area (s) covered by a colloidal particle in a time (t) is determined by means of microphotographs, the areas concerned being determined by projecting

the photographs on to paper, and measuring the areas by means of a planimeter. Then, where k is Boltzmann's constant (R/N), s is given by the equation

$$s = \pi^{\frac{1}{3}} \left(\frac{tkT}{\eta}\right)^{\frac{2}{3}},$$

where T is the absolute temperature, and η the viscosity of the medium. Experiments were made with suspensions of gamboge in water, glycol and glycerol, and the authors remark that the method is suitable for the determination of the viscosity of liquids obtainable in small quantities only. In their second paper the authors give a correction to this expression, pointing out that the correct value for s is obtained by multiplying the expression just given by $f(x)$, where x stands for Dt/σ^2, D being the diffusion coefficient $(kT/6\pi\eta\sigma)$, and σ being the particle radius. The function $f(x)$ is

$$f(x) = \frac{4}{6^{\frac{2}{3}}} \left(\frac{x^{\frac{1}{3}}}{\log 1{\cdot}26x}\right) \left(1 + \frac{0{\cdot}422}{\log 1{\cdot}26x} - \frac{0{\cdot}467}{(\log 1{\cdot}26x)^2}\right).$$

Podszus[16] describes a method by which Brownian movements may be made visible to the unaided eye. Very fine aluminium platelets, of the order of one or two μ in diameter are suspended in benzene containing some rubber solution. The suspension is stable and of a dark brown colour. When illuminated by a ray of light, the suspension clearly shows movements which are presumed to be Brownian in origin.

W. Haller[17] has studied the orientation and deformation of dispersed particles in flowing liquids. If Brownian motion is neglected the orientation approaches a state in which the longest direction is parallel to the line of flow. If Brownian motion is important the orientation may be something less than 45°. It is noted that V_2O_5 sols check this behaviour. If we are dealing with isotropic non-rigid particles a deformation is produced by the flow, and at high velocities and deformations the orientation may approach 90°.

E. Kappler[18, 19] has investigated the Brownian motion of a torsion balance and its variation with time and has deduced from his observations a very self-consistent value for Avogadro's number. He employed for this purpose a quartz fibre about 1 cm. long and about $0{\cdot}1\mu$ in diameter to the end of which was fixed a light mirror about 1 or 2 sq. mm. in area. Light, after passing through a small orifice was concentrated on the mirror and reflected to a camera. As an alternative to the photographic method direct observation of the movement of the mirror was attempted. Of nine experiments, six were made at atmospheric pressure, two at a pressure of 19 mm. of mercury, and one at a pressure of 50 mm. These gave respectively for Avogadro's number divided by 10^{22}, 60·26, 60·92, 59·55, 60·00, 60·62, 61·06–62·65, 59·74–60·42. The mean of these figures gives $60{\cdot}57 \times 10^{22}$ for Avogadro's number.

In such a simply suspended system, which is completely characterized by a single coordinate ϕ, Avogadro's number is obtained from the simple equation

$$A\bar{\phi}^2 = RT/N,$$

where A is the torsion constant of the quartz fibre. Obviously the question rises whether Brownian motion does not set a limit to the use of measuring instruments

which depend on the motion of a suspended system. This matter has been discussed in detail in an important paper by Barnes and Silverman[20], who, considering a system such as that just described, point out that there is "a definite limit set by Brownian motion. However, as may readily be seen, the sensitivity of the system may be increased by diminishing A. For, whereas the Brownian movement fluctuation is represented by $A\dot{\phi}^3$, the constant turning movement produced by the quantity being measured is represented by $A\dot{\phi}$. The Brownian motion is then proportional to $1/A^{\frac{1}{2}}$, while the real deflection produced by our given energy quantity is proportional to $1/A$. Our net gain is therefore inversely proportional to $\sqrt{A}$. Mechanical stability and the patience necessary to read long-period instruments are the limiting factors of sensitivity for an ideal instrument." The authors discuss in detail all important suspended systems—torsion pendulums, elastic rods and strings, vibrating membranes, chemical balances, spring balances, galvanometers and electrometers. The systems concerned are of very varying degrees of mechanical complexity, but the principles involved are sufficiently illustrated by the study of the simple quartz-thread balance investigated by Gerlach and by Kappler.

Wells and Lange[21] have described a very simple device for demonstrating Brownian movement in gases. Three parallel holes A, B and C ($\frac{5}{16}$ in. in diameter, $\frac{1}{2}$ in. deep) are drilled in a brass rod. A hole $\frac{1}{8}$ in. in diameter, perpendicular to the diameter of the rod, is drilled through A and B into C. A thin brass tube, carrying two narrow horizontal slits fits into A so that the slits are coincident with the ends of the $\frac{1}{8}$ in. hole. An intense beam of light is passed through the horizontal hole, the slits in A acting as a sort of collimator, and illuminates smoke particles introduced into B. The holes A, B and C, the axes of which are vertical, are closed above by a glass plate, and are painted inside dead black. The whole apparatus may be placed on a microscope stage, and the Brownian motion may be seen very clearly using a magnification of about one hundred.

Finally we may note the work of Whitmore, Richmond and McTaggart[22] who have carried out an interesting series of experiments on thin films on a water surface. If, following the usual procedure, a little oil is dissolved in a volatile solvent, and a drop of the solution is placed on a clean water surface, it spreads over the surface leaving, after evaporation of the solvent, a thin film of oil on the water surface, which film may, in its behaviour, show close analogies with the state of a gas, a liquid or a solid. Suppose now the film to be compressed to such a stage as to give a monomolecular layer. With some substances, further compression may result in the development of a "myriad of little heaps, visible in the microscope, where the surplus oil collects leaving the remaining area apparently monomolecular and optically empty. These little heaps, presumably lens-like or globular in form, have frequently a vigorous Brownian motion, which at once attracts attention. The formation of these small masses of oil as well as their movements must be due to forces between the molecules of the film and of the water developed as the film is reduced in area."

In the present paper the authors have considered carefully the conditions under which these droplets exhibiting this so-called Brownian motion are formed. They

study the conditions under which liquid films of triolein and of oleic acid pass from the monomolecular state to the state showing Brownian motion and they find that

(i) the transition occurs at a lower pressure when the temperature is raised;

(ii) the films remain liquid in both states;

(iii) the volume of a droplet (probably lenticular in shape), is that of a sphere of diameter $0{\cdot}5\mu$.

The authors further find that films of palmitic and stearic acid show Brownian motion at low pressures but become rigid at higher pressures and that Brownian motion appears when a solid film of tripalmitin is heated to its melting-point. Benzyl benzoate behaves in a very peculiar fashion. If a drop is placed on water "a thin film spreads immediately leaving one or more lenses which show considerable agitation and burst into smaller lenses, this process repeating itself until no more lenses can be seen. During this time, the length of which may be several minutes, a constant pressure of about 3 dynes per centimetre is maintained. After the last visible trace of oil disappears, the pressure falls, at first rapidly and then more slowly, towards zero. On compressing the film a cloud of droplets in Brownian motion is obtained as in other cases of liquid oils. The amount of oil that can be put on the surface in this way is quite large compared with other oils."

§ 3. SURFACE TENSION

During the last year or so, there have been considerable, if unsensational, additions to the literature. The present review picks out a few of the more important topics treated, and leaves a systematic consideration of the general aspect of surface films and of adsorption problems to a special review which will, it is hoped, appear in the third volume of these reports. For an admirable review of the main problems of surface chemistry now some two or three years old the reader is referred to Langmuir's article[1] on surface chemistry.

Several experimental investigations of the validity of Antonoff's rule have been published. The rule states that the interfacial tension between two partially miscible or non-miscible liquids is equal to the difference between the surface tensions of the saturated solutions. Loman and Zwikker[2], who subscribe to the views of Nellensteyn and Roodenberg[3] that the rule may be applied to solid-liquid interfaces, state the rule in the simple form that the interfacial solid-liquid tension is equal to the difference between the surface tensions of the liquid and of the solid. Adopting the authors' notation, we consider two liquids of surface tensions σ_1 and σ_2, having contact angles ϕ_1 and ϕ_2 with a certain solid whose surface tension is x. Then if one of the liquids be in contact with air and with a slab of the solid we obtain, either by resolving tensions, or by the more satisfactory method of applying the energy principle to a small displacement of the liquid over the solid,

$$x = \sigma_1 \cos \phi_1 + y_1.$$

Similarly for the other liquid

$$x = \sigma_2 \cos \phi_2 + y_2,$$

where y_1 and y_2 are the interfacial tensions of the liquid with respect to the solid.

Further, if Antonoff's rule holds in the form stated, we have

$$\sigma_1 = x + y_1; \quad \sigma_2 = x + y_2,$$

leading to the result

$$\frac{\cos \frac{1}{2}\phi_1}{\cos \frac{1}{2}\phi_2} = \sqrt{\frac{\sigma_2}{\sigma_1}} = \text{const.}$$

The authors point out that, the standard liquids used being water and mercury, if ϕ_w and ϕ_m be determined for drops of water and of mercury resting on *different* solids, a plot of $\cos \frac{1}{2}\phi_w$ against $\cos \frac{1}{2}\phi_m$ should result in a straight line passing through the origin and having a slope equal to $\sqrt{72{\cdot}8}/\sqrt{475} = 1/2{\cdot}554$.

The authors have designed a cathetometer suited for such measurements and, having regard to the difficulty of the observations and the doubt concerning the value of the surface tension of mercury, they regard Antonoff's rule in the form stated as holding very closely. An example which shows the consistency of their contact-angle measurements may be of interest.

Angle of contact with *Solid. Dark Silesian Granite*

(i) water: 140°·9, 140°·6, 140°·9, 141°·0, 140°·7—Mean 140°·8.

(ii) mercury: 57°·0, 54°·5, 57°·5, 56°·0, 55°·0—Mean 56°·0.

The value to be assumed for the surface tension of mercury is a matter of difficulty; the modern literature gives values ranging between 505 and 460 dynes per centimetre, and the authors take the value of the surface tension as 475 ± 25 dynes per centimetre. Hence, taking the surface tension of water at 20° C. as 72·8, the limiting slopes of the test lines should be taken as 2·621 and 2·486 (or the reciprocals of these numbers). The points obtained experimentally show a certain amount of scatter, but if we calculate the surface tension of mercury from the fact that the slope is equal to $\sqrt{\sigma_2}/\sqrt{\sigma_1}$, we find a value of 476 dynes per centimetre with a probable error of 1·6.

The validity of Antonoff's rule being regarded as established, the authors then proceed to calculate the surface tensions of solids from equations of the type

$$\kappa = \sigma_w \cos^2 \tfrac{1}{2}\phi_w, \quad \text{or} \quad x = \sigma_m \cos^2 \tfrac{1}{2}\phi_m.$$

Thus, Swedish red marble, which gave a contact angle of 146°·5 with mercury, and 80° with water, has a surface tension 42·7 dynes per centimetre as calculated from the water contact angle, and 39·6 calculated from the mercury contact angle, if the surface tension of mercury is taken as 475. It should be noted that $\kappa < \sigma_1 < \sigma_2$, so that "if we have to do with substances possessing a higher surface tension than water, only mercury measurements are available".

The authors draw attention to a curious slip made in another worker's determination of the surface tension of a solid[4]; the error has also been commented on by Adam[5]. In these experiments, a liquid paste is made having an angle of contact of 90° with glass, so that it neither rises nor falls in a tube of the glass in question. It was assumed that the surface tension of the glass was equal to that of the paste, which latter value was measured as 130 dyne-cm.$^{-1}$ The authors remark that if

$$\sigma \cos \phi = x - y, \quad \text{and} \quad y = \sigma - x,$$

we have

$$x=\sigma\cos^2\tfrac{1}{2}\phi \quad \text{and} \quad y=\sigma\sin^2\tfrac{1}{2}\phi,$$

so that when $\phi=90^\circ$

$$x=y=\tfrac{1}{2}\sigma.$$

Hence the surface tension of the glass in the experiments under consideration is about 65 dyne-cm.$^{-1}$ Actually, this value is low; rigorous cleaning of glass may raise its surface-tension from a value of 64 dyne-cm.$^{-1}$ to 120 dyne-cm.$^{-1}$

A careful investigation of the validity of Antonoff's rule as regards liquids has been carried out by E. G. Carter and D. C. Jones[6]. The authors have determined, by a capillary-tube method, the interfacial tensions against water of some twelve pure organic liquids of widely different composition. The authors stress a number of points of experimental interest—the necessity for rigorous purification of the substances under test, and the dangers attendant on the use of pyrex glass, for example. In all the experiments to determine interfacial tensions, the presence or absence of a finite contact angle was tested by measuring, with due precaution against errors of refraction, the width and height of the meniscus in the capillary. Pyrex glass gave contact angles at the water-liquid interface which, for methylene iodide, reached a value of 18°·5. Soda-glass was found to be far more satisfactory, in that it gave zero contact-angles for most of the systems under test. An interesting observation is made concerning the determination of the surface tension of the aqueous phase. The maximum bubble-pressure method was used, and it was found, for example, that the surface tension of water saturated with benzene *was practically that of pure water*, if it was determined by bubbling pure air through the apparatus. Special precautions were necessary to ensure that the air was completely saturated with benzene vapour and that the saturator and all the tubes connected therewith should be kept at the temperature at which the surface-tension observations were made.

The validity or non-validity of Antonoff's rule is closely connected with the value of the *spreading coefficient*. Thus if a film of liquid (L) on the surface of water (w) increases by unit area the change (S) in the free surface energy of the system is called the spreading coefficient and is given by

$$S=\sigma_w-(\sigma_L+\sigma_{wL}).$$

S must be positive for spreading to take place. Evidently, where W_1 is the work of adhesion of the liquid to water, and W_2 the work of cohesion of the liquid (see p. 9), we have

$$S=W_1-W_2.$$

We have to differentiate between the *initial* spreading coefficient between two pure liquids, and the *final* spreading coefficient between two mutually saturated liquids. Calling this latter coefficient S' we have

$$S'=\sigma_w'-(\sigma_L'+\sigma'_{wL}).$$

But, if Antonoff's rule holds,

$$\sigma'_{wL}=\sigma_w'-\sigma_L',$$

from which it follows at once that the final spreading coefficient S' must be zero.

This gives the conditions under which Antonoff's rule is valid.

The authors find confirmation of the rule for a number of systems, but in other systems find large deviations from the rule. "These latter can be divided into two groups. The first group contains systems where the organic substance has a negative initial spreading coefficient on water. The substances in the second group have high positive initial spreading coefficients on water, and it is the opinion of the authors that their non-conformity with this relationship may be due to an oriented adsorption film on the saturated aqueous phase."

Flotation processes play an increasingly important part in the treatment of suitable ores such as those containing the sulphide minerals of copper and the mixed sulphides of lead and zinc. The magnitude of the operations concerned may be gauged from the fact that a recent treatise gives costs for plant treating upwards of 4000 tons of ore per day, and the amount of ore treated annually reaches hundreds of thousands of tons. Broadly speaking, the standard procedure may be analysed under five heads: (*a*) crushing, (*b*) grinding, (*c*) flotation, (*d*) concentrate filtration and (*e*) tailing disposal. It is the examination of the principles involved in the actual flotation processes which is of fundamental interest to the student of surface forces; and, despite a rapidly increasing literature, there are many unsolved problems which will repay careful study.

The physical (or physico-chemical) problem involved is that of suspending a finely ground ore in water containing some substance capable of producing a stable froth. Moreover, the mineral which it is desired to concentrate must cling to the surface while the gangue remains in the body of the water, and a reagent must be added which produces this effect.

Obviously energy considerations play a large part in any theoretical deductions, and much useful information may be obtained from the use of the equation which has been named after Dupré[7]. Remembering that surface tension and free surface energy per unit area are equivalent terms, we see that in separating unit area of a solid (S) from a liquid (L), the solid-liquid interface disappears and two surfaces (solid-air and liquid-air) appear, and that what we may term the work of adhesion (W_{SL}) may be written

$$W_{SL}=\gamma_{SA}+\gamma_{LA}-\gamma_{SL},$$

where the symbol γ is used for surface tension. If the liquid has a contact angle θ with the solid, consideration of the change in potential energy involved in a small displacement leads to the well-known equation

$$\gamma_{SA}=\gamma_{SL}+\gamma_{LA}\cos\theta,$$

and the work of adhesion may be calculated from the equation

$$W_{SL}=\gamma_{LA}(1+\cos\theta),$$

involving quantities that may be directly measured. (The values for W_{SL} involved in putting $\theta=0°$, $90°$ or $180°$ should be noted.)

With fine particles, gravity considerations are secondary, and whether a liquid advances to cover a particle or whether the particle floats at the interface depends on the sign of the potential energy change. For example, as a simple figure shows,

if a liquid is in contact with a solid and with air, a displacement involving an *increase* of unit area between solid and air involves an *increase* in the potential energy of the system of

$$\gamma_{SA} - (\gamma_{SL} + \gamma_{LA} \cos \theta),$$

and the direction of the displacement is governed by the sign of this expression. It is evident that a knowledge of the magnitude of the contact angle and of the effect of different reagents in altering the contact angle is of great importance in flotation processes. Very valuable information concerning contact angles is to be found in a paper by N. K. Adam and G. Jessop[8], where is to be found an explanation, in terms of friction, of the difference between the contact angles of a liquid advancing or receding over a surface—a matter of some importance. Thus if F is a friction force (per unit length) which just prevents motion one way or the other, we may write

$$\gamma_{SA} - \gamma_{SL} = \gamma_{LA} \cos \theta_1 + F,$$

$$\gamma_{SA} - \gamma_{SL} = \gamma_{LA} \cos \theta_2 - F,$$

which combined with the equilibrium equation

$$\gamma_{SA} - \gamma_{SL} = \gamma_{LA} \cos \theta$$

gives

$$\cos \theta = \cos \theta_1 + \cos \theta_2.$$

Obviously, in considerations of flotation, a knowledge of contact angles is of primary importance and I. W. Wark[9] has considered in some detail the significance of the contact angle in flotation, in an investigation of the principles underlying the adhesion between a bubble of air and a solid. It is known that an air-bubble in water will *not* adhere to a clean surface of any of the common sulphur minerals or to metals. If however the surface of the mineral adsorbs a unimolecular film of a xanthate, the air of the bubble spreads to a limited extent over the surface of the mineral, partly replacing the aqueous phase. Equilibrium is determined by the contact-angle equation. The paper goes on to discuss the equation to the surface of a stationary bubble of air in water, develops equations connecting the volume of the bubble, the angle of contact, and the area of air-solid contact, and considers arithmetical solutions in the manner of Bashforth and Adams[10]. The consequences of Dupré's equation are then discussed, and the effect of hysteresis in measuring contact angles. The paper then proceeds to consider the stability of air-mineral attachments and criticizes the statement of Edser that "no particle could float stably but for the possibility of variation of the contact angle, for if this were constant a slight tilt would inevitably cause the particle to sink" and of Sulman that "but for hysteresis a mineralized bubble could have but a brief existence". The general argument is extended to moving systems and the maximum size of a particle which will float is also discussed.

The author in conjunction with various colleagues[13, 14, 15] has investigated a number of allied problems. Wark and Cox[15] investigate the nature of the adsorption of the soluble collectors, the relation between contact angle and the constitution of the collectors and criticize that theory of flotation developed by

Ostwald[16] in which, in the action of soluble collectors, only a ring of the collectors is supposed necessary for flotation, the location of the ring being the air-mineral-water line of contact, a theory in direct opposition to that advanced by the authors, that for soluble collectors a unimolecular film is adsorbed by the surface of the mineral, and that for oleaginous collectors a continuous thin film spreads over the surface.

The recovery by flotation of mineral particles of colloidal size is discussed by Gaudin and Malozemoff[17] who assert that non-flotation of colloidal sulphide mineral particles is "due to their being unable to come into contact with gas bubbles because of fine size and dispersion and that the flotation of colloidal minerals is facilitated by selective flocculation".

Bradley[18] attempts, by a very direct method of attack, to measure the cohesive force between solid surfaces and to estimate the surface energy of solids. In the experimental arrangement a small quartz sphere is suspended from a quartz spiral hanging vertically in a tube. A second sphere is attached to a vertical rod which may be raised or lowered by means of a cam at the lower end of the rod. The rod is raised until the spheres touch, and is then slowly lowered until contact is broken, when the elongation of the previously calibrated spring is measured. Experiments are repeated for spheres of different diameters. If in any one experiment, the diameter of one sphere is d_1, and that of the other d_2, it is found that a plot of the cohesive force against $d_1 d_2/(d_1+d_2)$ yields a straight line which passes through the origin.

On the assumption that the law of force is λ/x^n, the author calculates the mutual potential energy of two spheres and the attraction between them. He finds an expression for the attraction in the form

$$\frac{4\pi^2 q^2 \lambda}{(n-1)(n-3)(n-4)(n-5)}\left(\frac{r_1 r_2}{r_1+r_2}\right)\left(\frac{1}{d^{n-5}}\right),$$

where q is the number of attracting molecules per unit volume and d is the nearest distance between surface molecules of the spheres of radii r_1 and r_2.

The total surface energy is given by[19]

$$U=\frac{\pi q^2 \lambda}{(n-1)(n-3)(n-4)(n-5)}\left(\frac{1}{d^{n-5}}\right).$$

Hence, if the diameters of the spheres are d_1 and d_2, we have the cohesive force given by $A d_1 d_2/(d_1+d_2)$, where, as we have seen, A may be determined experimentally. The surface energy is equal to $A/2\pi$ and is consequently known. The author finds that the surface energy of quartz is 33·8 ergs per sq. cm.

Of other experimental investigations of capillary constants, mention may be made of the measurements of surface and interfacial tensions made by Speakman[20], who uses two capillary tubes joined as a U-tube and measures the difference of the pressures required to force the meniscus to the same level in each tube successively. If these pressures are P_1 and P_2 then $\Delta P\ (=P_2-P_1)$ is a measure of the surface tension. As a matter of experimental interest it may be noted that the author finds

it necessary to connect both sides of the apparatus to the manometer, and to insert on each side of the manometer large flasks to steady the pressure. He does not recommend alcohol as a manometric liquid inasmuch as it is too hygroscopic, increasing in density by as much as 1 or 2 per cent in a month. The author uses *cyclo*-hexane. Results are given for the surface tensions of benzene and water, and for the interfacial tensions of the systems water-benzene, water-ether and water-chloroform. In a later paper[21] the author describes experiments made to determine the surface tensions and partial vapour pressures of aqueous aniline solutions. The previous results of Worley[22] and of Seith[23] for the variation of surface tension with concentration for aqueous solutions of aniline are discordant; the author's values agree approximately with those of Seith. Values of the surface tension interpolated from a smooth curve were plotted against logarithms of similarly interpolated vapour pressure, and the adsorption was calculated from the slope $(-d\gamma/d.\log p_a)$ of this curve. The values at saturation correspond to a maximum adsorption of $4{\cdot}27 \times 10^{14}$ molecules per sq. cm. or an area per molecule of 23·4 sq. A., a more reasonable figure than the high value of 28·5 previously given[24].

An important contribution to the experimental side of the subject has recently been made by Dr W. N. Bond[25], who has supplied an entirely novel method for the determination of the surface tension of a freshly formed, and therefore uncontaminated surface. If two nozzles of suitable dimensions are placed in such a manner as to deliver two cylindrical jets of liquid with their axes vertical and collinear, the impact of the jets causes the development of a horizontal disk of liquid. The surface tension of the liquid is obtained from a determination of the maximum radius (R) of the liquid disk, the internal radius (r) of the nozzle, and the volume (Q) of liquid of density ρ, supplied in unit time. Devoid of all minor corrections, the surface tension γ is given by

$$\gamma = Q^2\rho/8\pi^2Rr^2.$$

Corrections to this formula are necessary on account of the weight of the liquid disk, the nett upward momentum supplied by the jets, the downward acceleration of the liquid as it passes round the curved edge of the liquid disk, and the effect of surface tension in increasing the pressure in the jets. These corrections are duly made, and, as the result of nine very consistent experiments made with two sets of nozzles ($2r = 0{\cdot}140$ and $0{\cdot}185$ cm. respectively) the author finds, for water at 15° C.

$$\gamma = 73{\cdot}83 \pm 0{\cdot}13 \text{ dyne-cm.}^{-1}$$

Schultze[26] has determined the surface tension of water from observation of the rise of the liquid between parallel glass plates fixed at a distance of b cm. apart by means of a specially designed precision apparatus for ensuring the parallelism of the plates. The plates were 50 mm. × 10 mm. and varied in thickness from 0·4 mm. to 2·0 mm. Schultze records that the capillary rise depends on the *thickness* of the plates, and that for a fixed value of b, the rise increases with plate thickness. Sata and Kurano[27] have also studied the capillary *plate* method for the measurement of surface tensions, and have investigated the effect of the finite width of the plate on capillary elevation.

Achmatov[28] describes a micro-method for the determination of the surface tension of a small quantity of liquid. Two capillary tubes of radii r and R are joined together and contain a thread of liquid. Evidently the thread may be kept in equilibrium by suitably inclining the tube, or by keeping the tube horizontal (or vertical) and applying a suitable known external pressure to one end of the tube. Simple equilibrium equations are easily written down.

Tammann and Boehme[29] heat strips of gold foil, and draw curves showing how $\Delta l/l$ varies with temperature for different velocities of heating. From the curves they determine the temperature at which wrinkling sets in, and by comparing these with the appropriate loading curves they deduce for the surface tension of gold:

Temp. (o° C.)	700°	750°	800°	850°
γ (mg./cm.)	1230	1220	1200	1180

There are a few theoretical points to record. Rocard[30] has applied general hydrodynamical theory to capillary problems, treating the subject on classical lines. He obtains a series of general expressions for P_{xx}, P_{xy}, P_{xz} in a dense fluid. Hence he deduces hydrodynamic equations, involving capillary terms, in the form

$$\rho\frac{du}{dt}+\frac{\partial P}{\partial x}+\mu\Delta u+\frac{\mu}{3}\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}\right)-\frac{a\sigma^2}{5M^2}\cdot\rho\cdot\frac{\partial}{\partial x}\cdot\Delta\rho=0.$$

The equations have applications in the theory of sound. The study of the equilibrium of a non-homogeneous fluid made with the aid of these equations leads to results which include a surface-tension term, the surface tension (γ) being given by

$$\gamma=\frac{a\sigma^2\lambda}{5M^2}\int\left(\frac{\partial\rho}{\partial x}\right)^2 dx.$$

Here M stands for molecular weight and σ for molecular diameter; a is the constant which appears in van der Waals' equation, and λ is a factor which depends on the law of force chosen, but does not differ greatly from unity. The length x is normal to the surfaces $\rho=$ constant, the integral being taken through the whole region where ρ varies. All the ordinary consequences of the existence of a surface tension—the relation between pressure and curvature, for example—follow at once.

Using the equation of state

$$P=RTv^2\Big/\left(v-\frac{b}{3}\right)-a/v^2,$$

the author finds that "pour des températures un peu éloignées de T_c"

$$\gamma=7{\cdot}32\sigma\sqrt{\lambda}\left(\frac{\rho_1-\rho_2}{\rho_c}\right)P_c\left(1-\frac{1{\cdot}03T}{T_c}\right).$$

He remarks that "les vérifications de la formule sont pratiquement parfaites pour les corps sans dipôle électrique (Hg, O_2, N_2, ...) ou à dipôle faible et à molécule à peu près sphérique (Cl_2, C_6H_6, CCl_4, NH_3). Il y a, par contre des écarts notables pour les corps à molécule très allongée et à fort dipôle permanent (eau, alcools, acetone, etc.)".

Various attempts have been made to find relations between the temperature coefficients of surface tension and of expansion. The problem has recently been attacked by Wheeler[31] who has attempted to develop a theory of liquids on the assumption that, at temperatures of low vapour pressure, the apparent diameters of the molecules of a liquid increase with temperature in such a way that the molecules on an average continue to touch. The author deduces expressions for the internal latent heat, the molecular heat, and the variation of surface tension with temperature at constant pressure. He draws the conclusion that for unassociated substances, the law of molecular attraction is that of the inverse 10th power of the distance. For surface tension in particular he finds that, where α is the volume coefficient of expansion at constant pressure, and p is the exponent in the law of attraction

$$-\frac{1}{\gamma}\left(\frac{\partial\gamma}{\partial T}\right)_p = \alpha.\frac{p+1}{3}.$$

He further deduces a general parachor law in the form

$$\gamma = KV^{-\frac{p+1}{3}},$$

which, if p is taken as 10, involves the exponent 11/3 instead of the usual 4.

The ratio of $-\frac{1}{\gamma}\left(\frac{\partial\gamma}{\partial T}\right)$ to α is therefore equal to $(p+1)/3$, and it is of interest to quote a few of the experimental figures:

Substance	T	$-\frac{1}{\gamma}\frac{\partial\gamma}{\partial T}$	α	$\frac{p+1}{3}$	p
$CHCl_3$	288	0·00500	0·0012	4·16	11
CCl_4	288	0·00435	0·0012	3·62	10
$(CH_3)_2CO$	283	0·00500	0·0014	3·58	10
$(C_2H_5)_2O$	273	0·00531	0·0015	3·54	10
C_6H_6	288	0·00447	0·0012	3·73	10
C_6H_{14}	283	0·00563	0·0014	4·01	11

The coefficient of expansion of an unassociated liquid, defined as $-\frac{1}{\rho}\frac{\partial\rho}{\partial\theta}$, may at once be deduced from an expression

$$\rho = 2\rho_c[(1-m)^{0\cdot3} + (1-0\cdot5m)]$$

developed some time ago by the present writer[32], which represents the variation of *orthobaric* density with reduced temperature (m) over the whole range of existence of an unassociated liquid. If further we assume the power law, in its simplest form

$$\gamma = \gamma_0(1-m)^{1\cdot2},$$

connecting surface tension with temperature, we can easily calculate the temperature coefficient of surface tension, and evaluate the ratio under consideration. It will be found that the ratio is a function of the reduced temperature, having a limiting value of 3 for a zero value of m, and that it does not increase very rapidly for fairly small values of m, its value for $m = 0\cdot5$ being between 3 and 4, which is in good agreement with the values obtained directly from the experimental figures as the table shows.

A number of formulae connecting surface tension and other physical constants with density have been put forward by Herz[33, 34, 35]. He finds that, for surface tension (γ), viscosity (η) and latent heat (L), the ratios

$$\frac{\gamma^{\frac{1}{3}}}{\rho_l-\rho_e}, \quad \frac{\eta^{\frac{1}{3}}}{\rho_l-\rho_e}, \quad \frac{L}{\rho_l-\rho_e}$$

are sensibly independent of the temperature. In the latest of these papers, he finds that

$$\rho_l-\rho_e=K(\rho_e-\rho_v)^{2\cdot5},$$

and hence that the ratios

$$\frac{L^{\frac{2}{5}}}{\rho_e-\rho_v}, \quad \frac{\eta^{\frac{2}{5}}}{\rho_e-\rho_v}, \quad \frac{\gamma^{\frac{2}{5}}}{\rho_e-\rho_v}$$

are sensibly constant.

The accuracy and implications of these deductions, as far as surface tension is concerned, may easily be followed out by an application of the power law connecting surface tension and temperature, and the density-temperature relation just mentioned[32], together with a similar relation for the density of the saturated vapour.

§ 4. RAYLEIGH'S PRINCIPLE

The late Lord Rayleigh enunciated a principle of great power and scope[1] for the determination of the frequency of the fundamental mode of vibration of an elastic system. *If an elastic system is vibrating in its fundamental mode the kinetic and potential energies are so distributed as to make the frequency a minimum.*

Suppose, then, that it is difficult, or impossible, to calculate the period of this fundamental vibration. If we constrain the system to vibrate in a mode which approximates to the natural mode, but which is amenable to calculation, then, by the principle, the frequency so calculated cannot be less than the fundamental frequency of the natural vibration. Such a calculation—which is most conveniently made by an application of the energy principle—gives us an upper limit to the frequency.

How close is this limit? We can hardly say, until we know something about a lower limit. Rayleigh did not discuss this problem, but it may be noted (and it is characteristic of Rayleigh's instinct for the essentials of a physical problem) that the approximation thus calculated to the natural frequency is much closer than the approximation of the constrained mode to the true mode.

Thus, to take a simple example. A flexible string, stretched under a tension T, is clamped at its ends, and vibrates in its fundamental mode. Let the string be of length l and linear density σ. Then, if we assume its outline to be parabolic, say

$$y=cx\,(l-x),$$

and put, as usual, $\omega^2=4\pi^2n^2$, where n is the frequency, we have for the mean kinetic and potential energies

$$\overline{K}=\tfrac{1}{4}\omega^2\int_0^l \sigma y^2 dx=\frac{\omega^2\sigma c^2 l^5}{120}$$

and $$\overline{P} = \tfrac{1}{4}T\int_0^l \left(\frac{dy}{dx}\right)^2 dx = \frac{Tc^2l^3}{12},$$

and since $\overline{K} = \overline{P}$, we find $$\omega^2 = 4\pi^2n^2 = \frac{10T}{\sigma l^2},$$

the true value being, of course, $\pi^2T/\sigma l^2$, so that the simple approximation involves an error of about 0·6 per cent in the frequency[2].

Recently, Prof. G. Temple[3] has discussed very exhaustively the problem of finding lower limits to the true values, and so that problems amenable to this mode of attack may be solved "with close and known degrees of approximation". Prof. Temple and Dr Bickley have described the method in detail, and have given many applications, both to vibrating systems and to equilibrium configurations, in a recently published volume[4].

§ 5. VISCOSITY

The most significant advances which have been made during the last few years are concerned with theories of liquid viscosity, and with the development of theoretical or semi-empirical relations connecting viscosity and other physical constants. Of these, the most important is the relation connecting viscosity and temperature. The older equation,

$$\eta = C/(A+t)^n,$$

proposed fifty years ago by Slotte[1] has had a considerable vogue, and, indeed, represents with considerable accuracy the behaviour of a large number of liquids over the range freezing- to boiling-point. But this, and most of the other formulae that have been put forward during the last generation, have been replaced by the simple and elegant formula

$$\eta = A \exp(b/T)$$

or

$$\log \eta = a + b/T$$

which is usually, and properly, known as Andrade's formula. As has been pointed out by Andrade and others, the formula has had a chequered history. Apparently its first mention occurs[2] in the Spanish literature of the subject in 1913; Kendall noted it in a footnote to a paper[3] published in the *Journal of the American Chemical Society* in 1917; in 1926, Dunn proposed it in a development of a simple kinetic theory of viscosity[4]; and in 1930 notes putting forward the formula were published in *Nature* by Andrade[5], and by Sheppard[6].

A number of isolated attempts to develop a theory of the viscosity of liquids have been made in recent years. That of Dunn, just mentioned, gives semi-quantitative reasons for assuming that the diffusion constant (D) of a liquid may be written

$$D = RT \exp(-Q/RT).$$

Hence the alternative expression for D in terms of Stokes' law,

$$D = RT/6\pi\eta r N,$$

leads to the required relation between viscosity and temperature.

Lederer[7] derives thermodynamically the relation

$$\log \eta = \frac{q_0}{RT} - 2{\cdot}75 \log T + ET + C,$$

q_0 being the heat of association at 0°, and E is equal to $\epsilon/4{\cdot}57$, where ϵ is a temperature coefficient in the equation

$$q = q_0 + T - \epsilon T^2.$$

Sato[8] has discussed a kinetic theory of liquids in the manner of Debye and Huckel. Dunn remarks, with some show of justification, that his final expression for η is "far from tractable, and is perhaps not likely to commend itself to many chemists on that account".

Shirokov[9] has developed a theory of viscosity for gases and liquids in terms of a van der Waals concept of the fluid state which assumes the form

$$\eta = \alpha \left(p + \frac{a}{v^2} \right) [v^2/(v-b)],$$

an equation which shows obvious affinities with Batchinski's equation[10] for liquids

$$\eta = C/(v-w).$$

Madge[11] has discussed a kinetic treatment, based on Maxwell's notion of a relaxation time. The discussion leads to a temperature equation of the type

$$\eta = A \exp(-CT)/(T-b).$$

Silverman[12] remarks that Maxwell's equation $\eta = Et$, where E is an elastic modulus and t a relaxation time "suggests that there exists in the liquid some structure, which breaks down and builds up periodically, capable of having elastic properties". On the view that the relaxation time is the average time of existence of such an associated group, the author arrives at the expression

$$\eta = KV \exp(Q/RT)/\beta \sqrt{T}$$

for the effect of temperature on viscosity. He derives also expressions for the ratio of the viscosity under a pressure P to the viscosity under atmospheric pressure.

Undoubtedly the most promising of these attempts to formulate a theory of liquid viscosity is that which we owe to Prof. Andrade[13, 14] who proposes a theory developed on classical lines which gives values for the viscosity of a simple liquid near its freezing-point in terms of other experimental constants, and from which may be derived the exponential formula connecting viscosity and temperature.

As the author points out, the mechanism of liquid viscosity is evidently different from that of gaseous viscosity—the different behaviour with respect to temperature is a sufficient pointer. He discusses in some detail the attack of J. D. van der Waals, junior, who, after arriving at the conclusion that the most important factor in the problem is the forces of collision, finds that

$$\eta \alpha \sqrt{T} \exp(-\epsilon/RT),$$

where the exponential term is of little importance, and of Brillouin who develops the notion that the impact of molecules gives rise to elastic waves which disturb neighbouring molecules. He finds that the viscosity is a function of the specific

volume and independent of the temperature. The author certainly does not overstate the case when he remarks that "in general, lack of agreement with experiment is the chief reproach that can be levelled against the older theories". The author believes that the theory of liquids should be approached, not from the side of gas theory, but from that of the solid state. He assumes that, in a system moving in plane parallel layers and possessing a normal velocity gradient, viscosity is due to a communication of momentum from layer to layer, and that this communication is brought about "by the temporary union at the periphery of molecules in adjacent layers, due to their large amplitude of vibration".

The development of the theory is based on two assumptions—that the arrangement of molecules in the liquid state is not fundamentally dissimilar to the solid-state arrangement, the difference between the two states being brought about by the large amplitude of the motions in the liquid state. Hence there is "for a liquid molecule a vibration about a slowly displaced equilibrium position, the frequency being the same as that for the solid state". The further assumption is made that when at every extreme libration "the molecules of one layer come into contact with the molecules of another layer they enter into a brief temporary union of short duration which enables them to attain a common velocity of translation". In terms of the frequency ν and the distance σ between neighbouring molecular centres the author finds that

$$\eta = \frac{4}{3}\frac{\nu m}{\sigma},$$

and, introducing the assumption that at the melting-point the frequency of vibration is that in the solid state, so that Lindemann's formula may be used, he finds for the viscosity at the melting-point

$$\eta = 5{\cdot}1 \times 10^{-4}\,(AT_m)^{\frac{1}{2}}/V^{\frac{2}{3}},$$

where A is the atomic weight and V the gram-atomic volume at the melting-point T_m. The formula gives values in very satisfactory agreement with the observed values as is shown in Table 1.

Table 1

Substance	Hg	Pb	Sn	Cu	Sb	Bi
η (calculated)	0·021	0·025	0·019	0·038	0·015	0·018
η (observed)	0·021	0·028	0·020	0·038	0·024	0·023

In a second paper, Andrade discusses the temperature variation of viscosity. "For transfer of momentum to take place the molecules considered must have a mutual potential energy $-E$, and the variation of viscosity with temperature will be governed by the fraction of molecules possessing this energy at extreme libration. According to Boltzmann the ratio of the number of molecules possessing energy $-E$ at a temperature T to the number possessing energy $-E$ at a temperature T' is

$$\exp\left[\frac{E}{k}\left(\frac{1}{T}-\frac{1}{T'}\right)\right].$$

This is the main factor in the temperature variation of viscosity and hence

$$\eta_T/\eta_{T'} = \exp(-E/kT')\exp(E/kT),$$

or
$$\eta_T = \eta_{T'}\exp(-E/kT')\exp(E/kT) = A\exp(C/T).$$"

It is worthy of note that this formula accounts for Porter's generalization[15] that if the viscosity of a liquid (I) at a temperature T_1 be η, and if for another liquid (II) the viscosity be η at a temperature T_2, then if different values of T_1 be taken and the corresponding values of T_2 found, the ratio T_1/T_2 is a linear function of T_1.

For,
$$A_1\exp(C_1/T_1) = A_2\exp(C_2/T_2).$$

That is,
$$\log A_1 + C_1/T_1 = \log A_2 + C_2/T_2.$$

Hence
$$C_2/T_2 - C_1/T_1 = \log A_1 - \log A_2 = K,$$

or
$$T_1/T_2 = \alpha + \beta T_1.$$

If account is taken of the variation of volume with temperature it is shown that as a result of very simple assumptions

$$\eta v^{\frac{1}{3}} = A\exp(C/vT).$$

This equation is applied to many substances over wide ranges of temperature, and fits the experimental results remarkably well.

The variation of viscosity with pressure is also discussed.

It is impossible to make more than passing mention of a number of other theoretical points which have recently been investigated. G. I. Taylor[16] has extended those investigations of Einstein which dealt with the viscosity of a liquid containing solid spheres in suspension to a liquid system in which the suspended particles are also liquid. The problem, which is intractable enough in general terms, is simplified by assuming that there is no slip at the surface of a drop in suspension in the liquid, that the tangential stress parallel to the surface is continuous at the surface of the drop, and that the surface-tension forces are sufficient to keep the drops nearly spherical. In these circumstances the author finds for the mean viscosity η_m of the liquid,

$$\eta_m = \eta\left(1 + 2{\cdot}5\Phi\frac{\eta' + \frac{2}{5}\eta}{\eta' + \eta}\right),$$

where η' is the viscosity of the liquid composing the droplets, and Φ is the fraction of the whole volume occupied by the spherical droplets. If η' is made infinite, we recover Einstein's equation

$$\eta_m = \eta(1 + 2{\cdot}5\Phi).$$

Prescott[17], in a paper dealing with the general equations of motion of a viscous fluid, remarks that some writers on the subject borrow the fundamental equations of elasticity and then deduce those of viscous flow by an analogy which they do not prove to exist. Other writers produce satisfactory proof with rather a greater call on mathematical analysis than is necessary, and others use methods still more doubtful than those who press the analogy with elasticity. In the present paper, the author deduces the equations of motion of a viscous fluid in the form

$$\frac{\partial p_1}{\partial x} - \mu\left(\nabla^2 u + \frac{1}{3}\frac{\partial \Delta}{\partial x}\right) = -\rho(a_x - X),$$

and two similar, where the symbols have their obvious meaning. He uses very simple and straightforward mathematical methods, under the assumption of laminar flow, and the proportionality of frictional drag to velocity gradient.

Of papers dealing with minor points of theory that by Francis[18] may be cited. The paper deals with the wall-effect in falling-sphere viscometry, and discusses the limitation of the well-known equation

$$\eta = \frac{2gR^2(\rho_1 - \rho_2)}{9u\,(1 + 2{\cdot}4\,R/r)\,(1 + 3{\cdot}3\,R/h)}.$$

Where the end-effect—the effect which is compensated by the introduction of the ratio R/h—is negligible, Francis proposes as a working equation

$$\eta = \frac{2gR^2(\rho_1 - \rho_2)(1 - R/r)^{2{\cdot}25}}{9u}.$$

He finds that this equation agrees well with experiment up to a ratio (for R/r) of 0·83, the limit of agreement of Ladenberg's formula being reached at a ratio of 0·09.

Reiner[19] has pointed out that the influence of a solid wall on the state of aggregation of a liquid may be investigated by means of viscosity measurements. If, for example, the viscosity of a boundary layer near the wall of a capillary tube differs from that of a liquid in bulk, it is possible, by means of viscosity measurements made at different pressures in several capillaries, to calculate the thickness of the boundary layer, its viscosity and the viscosity of the liquid as a whole.

Silverman and Roseveare[20] have put forward a very simple equation connecting viscosity and surface tension. Assuming the validity of Macleod's equation

$$\gamma = C\,(\rho_l - \rho_v)^4,$$

and of the equation proposed by Batchinski[10],

$$\eta\,(v - b) = C',$$

where v stands for specific volume, we have at once by elimination of the liquid density between these equations

$$\gamma^{-\frac{1}{4}} = A\eta^{-1} + B,$$

if we restrict ourselves to temperature ranges such that the vapour density ρ_v is negligible in comparison with the liquid density ρ_l.

It may be mentioned, in passing, that elimination of the temperature between Andrade's formula and the power law connecting surface tension and temperature leads to a simple semi-logarithmic relation between surface tension and viscosity at the same temperatures.

Turning now to the consideration of one or two experimental investigations, we find that Michels and Gibson[21] have carried out a set of experiments of fundamental importance on the variation of the viscosity of gases with pressure, carrying the measurements up to pressures of the order of a thousand atmospheres. In view of the different behaviour of gases and liquids, it was deemed of special interest to follow the changes in viscosity of a gas until its density became comparable with the density values in the liquid phase. Very little attention has been given to this aspect of the subject, and the authors are breaking practically new ground. They

use a Poiseuille method depending on the transpiration of a gas through a capillary tube. The viscometer is contained in a steel vessel with mercury at the bottom, the remainder being filled with oil. Pressure applied to the oil causes mercury to rise in the viscometer and to pass certain platinum contacts sealed therein, the times of passage being noted. A number of special corrections are necessary, and it is obvious that a knowledge of the compressibility isotherms of the gas is demanded.

Measurements were made up to pressures of a thousand atmospheres on nitrogen at 25°, 50° and 75° C. The viscosity-pressure curves show that at low pressures $(\partial\eta/\partial T)_p > 0$, and that at high pressures $(\partial\eta/\partial T)_p < 0$.

If viscosity be plotted against density, curves are obtained for the three different temperatures which do not cross as do the $\eta - p$ curves. $(\partial\eta/\partial T)_\rho$ is positive and approximately constant over the whole density range.

Enskog[22] has developed a theory of the viscosity of compressed gases which shows that a plot of η/ρ against ρ passes through a minimum, the relation between these variables being

$$\eta/\rho = \frac{1}{2{\cdot}545}(\eta/\rho)_{\min.}\left[\frac{1}{b\rho K} + 0{\cdot}800 + 0{\cdot}7614\, b\rho K\right],$$

$b\rho K$ being obtained from an equation of state

$$p + a\rho^2 = \frac{RT}{M}\rho(1 + b\rho K).$$

The author's experimental curves show this minimum unmistakably and give very good agreement with the values of η calculated from the equation. Enskog's theory leads to an equation

$$S = \frac{1}{0{\cdot}955}\sqrt{\left(\frac{M}{RT}\right)\left(\frac{\eta}{\rho}\right)_{\min.}},$$

for the diameter S of the gas molecule.

Substitution of the experimental values gives for nitrogen,

$$S = 2{\cdot}97 \times 10^{-8} \text{ cm. at } 25° \text{ C.}$$
$$S = 2{\cdot}94 \times 10^{-8} \text{ cm. at } 50° \text{ C.}$$
$$S = 2{\cdot}92 \times 10^{-8} \text{ cm. at } 75° \text{ C.}$$

During recent years, considerable attention has been given to the question of the relation between viscosity and concentration for solutions of strong electrolytes, and a paper by Lawrence and Wolfenden[23] discusses the subject critically, and describes further experiments with strong electrolytes. An equation of the form

$$\eta = 1 + A\sqrt{c} + Bc$$

has been assumed to give the relation between viscosity and concentration, c being the concentration in mols per litre, and η the *relative* viscosity. The constant A has been evaluated by Falkenhagen and his colleagues in terms of the ion-atmosphere theory. Falkenhagen and Vernon[24] give a convenient summary and papers by Jones and Talley[25], Joy and Wolfenden[26] and Cox and Wolfenden[27] describe

tests of the fundamental equation between η and c and of the agreement between the theoretical and observed values of the constant A. The value of the constant A may be written, for a uni-univalent salt as

$$A = \frac{1{\cdot}45}{\eta_0 \sqrt{(2D_0T)}} \phi\,(l_1, l_2),$$

where ϕ is a known, and simple, function of the equivalent ionic conductances at zero concentration, and η_0 and D_0 are the viscosity and dielectric constant of water at $T°$ C.

The paper by Jones and Talley contains a valuable historical résumé of the problem and describes a series of experiments to determine the relation between viscosity and concentration for solutions of KCl, $KClO_3$, $KBrO_3$, KNO_3, NH_4Cl, $CsNO_3$, sucrose and urea. The experiments were carried out in a viscometer of the Ostwald type, the times of flow being registered on a chronograph by means of a photo-electric cell. The results were tested by plotting $(\eta-1)/\sqrt{c}$ against $\sqrt{c}$, the slope and intercept of the resulting linear graph serving to determine the constants A and B. Table 2 shows the values of the constants so determined, and the order of agreement with the theoretical values.

Table 2

Solute	KCl	$KClO_3$	$KBrO_3$	KNO_3	NH_4Cl	$CsNO_3$	Sucrose	Urea
B (observed)	−0·0140	−0·0309	−0·0008	−0·0531	−0·0144	−0·092	+0·8786	+0·0378
A (observed)	0·0052	0·0050	0·0058	0·0050	0·0057	0·0043	0·000	0·000
A (theoretical)	0·0050	0·0055	0·0058	0·0052	0·0050	0·0051	0	0

Cox and Wolfenden show that the coefficient B is made up additively of two independent components characteristic of the anion and of the cation, and the investigation by Lawrence and Wolfenden is designed to determine whether this result is rigidly additive or merely a first approximation. To test this point (relative) viscosity-concentration measurements were carried out on several pairs of lithium and potassium salts with common anions. In the experimental work Washburn-Williams[28] viscometers were used and the "abstraction and addition" technique employed—that is, after measuring the time of flow for water, a definite small volume of water was removed from the viscometer by means of a special pipette, and replaced by an exactly equal volume of a solution of known concentration. Times of flow, which were of the order of 1000 sec. were determined with an error of less than 0·1 sec. Relative viscosities were calculated from the formula

$$\eta = \frac{t_s}{t_w}\,\frac{d_s - d_{\text{air}}}{d_w - d_{\text{air}}},$$

where the t's refer to times, and the d's to densities. As usual $(\eta-1)/\sqrt{c}$ was plotted against $\sqrt{c}$, and satisfactory linear plots resulted. The values of A determined from the graphs were in fair agreement with the Falkenhagen formula, and it was concluded that "the additivity of B coefficients for individual ions is probably no more than a satisfactory first approximation" and that "the principle of the additivity

of B breaks down unmistakably for the salts containing the tetraethyl-ammonium and the picrate ion, perhaps owing to ion-pair formation".

Pichot[29] has considered at some length what are known as *viscosity anomalies*. It is well known that, at low pressures, the rate of flow, Q, for certain colloidal solutions does not follow Poiseuille's law, and it is usually assumed, following Wo. Ostwald, that the flow may be expressed as

$$Q = kp^n,$$

where the index n is greater than unity. Such solutions are said to exhibit viscosity anomalies. If the flow of a liquid between two parallel planes of great extent be studied for pure water and for gelatine solution, a striking difference is shown. The velocities of different layers may be compared by suspending small particles of aluminium in the liquid and photographing these by intermittent light. It is found that with water a plot of velocity against distance from one wall results in a parabolic curve. The gelatine solution shows a flat head to the curve, indicating that a portion of the solution flows "as a plug". If one regards the liquid as flowing through a cylinder, then, besides the usual force on an element dS given by

$$dF_1 = \eta\,(dv/dr)\,dS,$$

we have tangential forces due to friction between the colloidal particles following the law

$$dF_2 = a\,dS.$$

Hence the steady-state equation is obtained by equating the forces arising from the pressure differences on an unit cylinder of radius r to

$$-\eta\,2\pi r\,dv/dr + a\,2\pi r.$$

The author has investigated this flow experimentally and has studied the magnitudes of the coefficients η and a.

The viscosity of heavy water has engaged the attention of a number of experimenters. Selwood and Frost[30] have measured a few characteristic properties of heavy water for various concentrations up to 92 per cent and have used these values for extrapolation to 100 per cent. In estimating the proportion of H^2 from the density measurements the value $d^{\text{v}}_{20} = 1{\cdot}1056$ was used for the density of pure heavy water.

Table 3 shows some of their results.

Table 3

Percentage heavy water	0	31	63·5	92	100
Density d^{v}_{20}	0·9982	1·0314	1·0664	1·0970	1·1056
$\mu = n^{\text{D}}_{20}$	1·33293	1·33138	1·32992	1·32849	1·3281
$\mu = n^{\text{C}}_{20}$	1·33094	1·32959	1·32824	1·32683	1·3265
η_{20} (millipoises)	10·87	11·4	12·7	13·7	14·2
γ_{20} (dyne-cm.$^{-1}$)	72·75	71·5	69·8	68·1	67·8

Lewis and Macdonald[31] have determined the viscosity of heavy water by measuring the time required to fill a pipette through a fine capillary under a known

pressure head. At each temperature the heavy water was directly compared with ordinary pure water. The authors find a very close approximation to linearity between viscosity and isotopic composition. Thus if we use the notation

$$\frac{\eta_2}{\eta_1} = \frac{\eta \text{ for } H^2H^2O}{\eta \text{ for } H^1H^1O},$$

then η_2/η_1 at 20° was found to be 1·249 by extrapolation from a 90 per cent sample, and 1·244 by extrapolation from a 70 per cent sample. The authors give the following table, in which η_2 is obtained by multiplying together the observed figures for η_1 and for the ratio η_2/η_1. Viscosities are given in millipoises.

Table 4

Temp. (o° C.)	5	10	15	20	25	30	35
η_2/η_1	1·309	1·286	1·267	1·249	1·232	1·215	1·198
η_1	15·19	13·10	11·45	10·09	8·95	8·00	7·21
η_2	19·88	16·85	14·51	12·60	11·03	9·72	8·64

It will be observed that the values here given are not in agreement with the value for η_{20} obtained by Selwood and Frost.

Bingham and Spooner[32] note that from the fluidity data just given the association of deuterium oxide is much less than that of ordinary water—1·34 as compared with 2·50. The low association of D_2O leads one to expect chemical inertness, and it is to inertness rather than to any active poisonous character that the failure of seeds to germinate and of starch to react with enzyme must be ascribed.

REFERENCES

§ 2. *Brownian motion*

(1) Burton. *Colloidal Solutions*. London.
(2) Einstein. *Ann. Phys.*, Lpz. (4), **19**, 371 (1906).
(3) —— *The Brownian Movement* (1926). London: Methuen.
(4) Langevin. *C.R. Acad. Sci.*, Paris, **146**, 530 (1908).
(5) Uhlenbeck and Ornstein. *Phys. Rev.* **36**, 823 (1930).
(6) Ornstein. *Proc. Acad. Sci. Amst.* **21**, 96 (1919).
Fürth. *Z. Phys.* **2**, 244 (1920).
(7) Uhlenbeck and Goudsmit. *Phys. Rev.* **34**, 145 (1929).
(8) Gerlach, W. *Naturwissenschaften*, **15**, 15 (1927).
(9) Metadier. *C.R. Acad. Sci.*, Paris, **195**, 649 (1932).
(10) —— *C.R. Acad. Sci.*, Paris, **197**, 29 (1933).
(11) Sato. *Z. Phys.* **80**, 822 (1933).
(12) —— *Z. Phys.* **86**, 667 (1934).
(13) Ornstein and van Wijk. *Physica*, 's Grav., **1**, 235 (1934).
(14) Brumberg and Vavilov. *Z. Phys.* **73**, 7833 (1932).
(15) —— —— *Z. Phys.* **75**, 746 (1932).
(16) Podszus. *Kolloidzschr.* **62**, 347 (1933).
(17) Haller, W. *Kolloidzschr.* **61**, 126 (1932).
(18) Kappler, E. *Ann. Phys.*, Lpz., **11**, 233 (1931).
(19) —— *Ann. Phys.*, Lpz., **15**, 545 (1932).
(20) Barnes and Silverman. *Rev. mod. Phys.* **6**, 162 (1934).
(21) Wells and Lange. *Rev. sci. Instrum.* **3**, 474 (1932).
(22) Whitmore, Richmond and McTaggart. *Trans. roy. Soc. Can.* **27** (§ iii), 113 (1933).

§ 3. *Surface tension*

(1) LANGMUIR. *Chem. Rev.* **13**, 147 (1933).
(2) LOMAN and ZWIKKER. *Physica*, 's Grav., **1**, 1181 (1934).
(3) NELLENSTEYN and RODENBERG. *Chem. Weekbl.* **30**, 189 (1933).
(4) ANTONOFF. *Phil. Mag.* **1**, 1258 (1926).
(5) ADAM. *Phil. Mag.* **3**, 863 (1927).
(6) CARTER and JONES. *Trans. Faraday Soc.* **30**, 1027 (1934).
(7) DUPRÉ. *Théorie Mécanique de la Chaleur* (1869).
(8) ADAM, N. K. and JESSOP, G. *J. chem. Soc.* **127**, 1863 (1925).
(9) WARK, I. W. *J. phys. Chem.* **37**, 623 (1933).
(10) BASHFORTH and ADAMS. *An Attempt to Test the Theories of Capillary Action.* Cambridge.
(11) EDSER. *B.A. Report on Colloid Chemistry*, **4**, 263 (1922).
(12) SULMAN. *Trans. Inst. Min. Metall.*, Lond., **29**, 44 (1919) and *Third Empire Min. and Met. Congress* (1930).
(13) COX and WARK. *J. phys. Chem.* **37**, 797 (1933).
(14) WARK and WARK. *J. phys. Chem.* **37**, 805 (1933).
(15) WARK and COX. *J. phys. Chem.* **37**, 815 (1933).
(16) OSTWALD, WO. *Kolloidzschr.* **58**, 179 (1932); **60**, 324 (1932).
(17) GAUDIN and MALOZEMOFF. *J. phys. Chem.* **37**, 597 (1933).
(18) BRADLEY. *Phil. Mag.* (7), **13**, 853 (1932).
(19) —— *Phil. Mag.* (7), **11**, 846 (1931).
(20) SPEAKMAN. *J. chem. Soc.* p. 1449 (1933).
(21) —— *J. chem. Soc.* p. 776 (1935).
(22) WORLEY. *J. chem. Soc.* **105**, 263 (1914).
(23) SEITH. *Z. phys. Chem.* **117**, 265 (1925).
(24) REHBINDER and TAUBMANN. *Z. phys. Chem.* **147**, 201 (1930).
(25) BOND. *Proc. phys. Soc.* **47**, 549 (1935).
(26) SCHULTZE. *Kolloidzschr.* **57**, 277 (1931).
(27) SATA and KURANO. *Kolloidzschr.* **65**, 191 (1933).
(28) ACHMATOV. *Kolloidzschr.* **66**, 266 (1934).
(29) TAMMANN and BOEHME. *Ann. Phys.*, **12**, 820 (1931).
(30) ROCARD. *C.R. Acad. Sci.*, Paris, **197**, 122 (1933).
(31) WHEELER. *Indian J. Phys.* **8**, 52 (1934).
(32) FERGUSON and MILLER. *Proc. phys. Soc.* **46**, 140 (1934).
(33) HERZ. *Z. anorg. Chem.* **159**, 304 (1927).
(34) —— *Z. anorg. Chem.* **175**, 278 (1928).
(35) —— *Z. anorg. Chem.* **203**, 271 (1933).

§ 4. *Rayleigh's principle*

(1) RAYLEIGH. *Theory of Sound*, **1**, 109 *sqq.* (1894).
(2) —— *Theory of Sound*, **1**, 112 (1894).
TEMPLE and BICKLEY. *Proc. roy. Soc.* **119**, 95 (1928).
(3) TEMPLE. *Proc. roy. Soc.* **119**, 276 (1928).
(4) TEMPLE and BICKLEY. *Rayleigh's Principle* (1933). Oxford.

§ 5. *Viscosity*

(1) SLOTTE. *Ann. Phys.*, Lpz., **20**, 257, 557 (1883).
(2) GUZMAN. *Annales de la Sociedad Española de Física y Química*, **11**, 353 (1913).
(3) KENDALL. *J. Amer. chem. Soc.* **39**, 1799 (1917).
(4) DUNN. *Trans. Faraday Soc.* **22**, 401 (1926).
(5) ANDRADE. *Nature*, Lond., **125**, 309 (March 1, 1930).
(6) SHEPPARD. *Nature*, Lond., **125**, 489 (March 29, 1930).

(7) LEDERER. *Kolloidchem. Beih.* **34**, 270 (1931).
(8) SATO. *Sci. Rep. Tôhoku Univ.*, **14**, 403 (1925).
(9) SHIROKOV. *Chem. Abs.* p. 457 (1932); *J. phys. Chem.* (*U.S.S.R.*), **3**, 174 (1932).
(10) BATCHINSKI. *Z. phys. Chem.* **84**, 643 (1913).
(11) MADGE. *J. phys. Chem.* **34**, 1599 (1930).
(12) SILVERMAN. *Trans. Faraday Soc.* **29**, 1285 (1933).
(13) ANDRADE. *Phil. Mag.* **16**, 497 (1934).
(14) —— *Phil. Mag.* **16**, 698 (1934).
(15) PORTER. *Phil. Mag.* **33**, 458 (1912).
(16) TAYLOR. *Proc. roy. Soc.* **138**, 41 (1932).
(17) PRESCOTT. *Phil. Mag.* **13**, 615 (1932).
(18) FRANCIS. *Physics*, **4**, 403 (1933).
(19) REINER. *Phys. Z.* **33**, 499 (1932).
(20) SILVERMAN and ROSEVEARE. *J. Amer. chem. Soc.* **54**, 4460 (1932).
(21) MICHELS and GIBSON. *Proc. roy. Soc.* **134**, 288 (1931).
(22) ENSKOG. *K. svenska VetenskAkad. Handl.* **63**, No. 4 (1922).
(23) LAWRENCE and WOLFENDEN. *J. chem. Soc.* p. 1144 (1934).
(24) FALKENHAGEN and VERNON. *Phil. Mag.* **14**, 537 (1932).
(25) JONES and TALLEY. *J. Amer. chem. Soc.* **55**, 624, 4124 (1933).
(26) JOY and WOLFENDEN. *Proc. roy. Soc.* **134**, 413 (1931).
(27) COX and WOLFENDEN. *Proc. roy. Soc.* **145**, 475 (1934).
(28) WASHBURN and WILLIAMS. *J. Amer. chem. Soc.* **35**, 737 (1913).
(29) PICHOT. *Physics*, **5**, 200 (1934).
(30) SELWOOD and FROST. *J. Amer. chem. Soc.* **55**, 4335 (1933).
(31) LEWIS and MACDONALD. *J. Amer. chem. Soc.* **55**, 4730 (1933).
(32) BINGHAM and SPOONER. *J. chem. Phys.* **2**, 107 (1934).

THE QUANTUM THEORY

Edited by Prof. G. Temple, D.Sc., Ph.D., F.Inst.P.

Part 1. QUANTUM THEORY OF ATOMIC NUCLEI

By R. Peierls, University of Manchester

§ 1. INTRODUCTION

Both the experimental data on nuclear structure and the theoretical methods of describing the properties of nuclei are in a process of rapid development and the present state of knowledge is probably very short of what we shall know in a few years. A report on this subject can therefore be only of a very temporary value. As, however, the large amount of data collected and of theoretical assumptions discussed make it very difficult to follow the development in detail, a short review of the present situation may be justified.

Since the discovery of the neutron by Chadwick and Curie-Joliot in 1932 it is generally assumed[1] that nuclei consist of protons and neutrons and that these particles may be treated like elementary particles, so that the only variables describing the state of the nucleus are the positions and the velocities of the protons and neutrons and the directions of their spins. Both particles are assumed to have an intrinsic angular momentum (spin) of $\frac{1}{2}$ unit ($=\frac{1}{2}h/2\pi$), where h is Planck's constant. For the proton this was proved by direct experiment, while for the neutron it is necessary in order to account for the observed angular momenta of nuclei.

Both these particles are heavy, the proton having about 1840 times the mass of the electron, while the mass of the neutron has almost the same value (the ratio of the two masses is 1·001, according to the latest determinations[2]). Since the energies of particles inside the nucleus are in general smaller than say 20 million electron-volts, their velocities will be smaller than about 20 per cent of the velocity of light, and therefore the refinements required by the theory of relativity can for most purposes be neglected. The kinetic energy of each particle may then be assumed to have the classical form $\frac{1}{2}Mv^2$ (M = mass, v = velocity) and the total energy of a nucleus will be equal to the sum of the kinetic energies of all particles and their interaction energy.

§ 2. INTERACTION FORCES IN THE NUCLEUS

Of the interaction energy, however, less is known; the simplest idea would be to expect that it is a sum of terms each referring to a pair of particles. This holds, for instance, for the interaction of electrons in an atom and, in a certain approximation, also for the interaction between atoms in a polyatomic molecule. Already in this latter case, however, the assumption is not rigorously true (i.e. the

interaction force between two atoms depends on the position of a third atom) and there is no reason why it should be generally true in the nucleus. It is, however, the simplest possible assumption and is used as a working hypothesis until the experimental evidence or theoretical arguments force us to abandon it.

Taking this assumption for granted, there are three types of interaction force to be considered: (1) The interaction between two protons, (2) between two neutrons, (3) between a proton and a neutron. The most interesting of these forces is the last, the proton-neutron interaction.

Three assumptions have been discussed for this force: (3*a*) Wigner[3] assumed it to be an attractive force of the conventional type, extending over distances of the order of a certain length a, so that the potential would vanish rapidly at distances greater than a and would approach a constant value B, if the distance approached zero.

(3*b*) Heisenberg[4] discussed another assumption. According to Heisenberg, there is no force in the ordinary sense of the word between the particles, but a tendency to exchange their positions. If in a certain moment the proton is at a point P and the neutron at a point P' so that the distance between P and P' is not too large, they would be very likely to change places and after a short time one would find the proton at P' and the neutron at P. Thus between a neutron and a proton at small distances a continuous interchange would take place and there would be an energy connected with this process. The possibility of such interchange phenomenon is peculiar to quantum mechanics and it is, in principle, the same fact which produces the valency forces binding the atoms of a homopolar molecule. For instance, in the hydrogen molecule ion which consists of a proton and a hydrogen atom the electron of the latter may jump over to the proton thus converting the proton into an atom and *vice versa*. The possibility of this transition lowers the energy of the two bodies as compared with a hydrogen atom and a proton at large distance, and thus produces an attraction. While in the case of the hydrogen molecule ion the possibility of an exchange is no new assumption but the result of the application of wave mechanics to the motion of the electron, Heisenberg *assumed* by analogy a similar type of exchange force between a proton and a neutron. This assumption, however, does not necessarily imply that the neutron (like the hydrogen atom) contains a proton and an electron; at the moment the question of the origin of this exchange force shall remain open and we shall only discuss to what extent such an assumption could describe the structure of nuclei. It has yet to be specified what happens to the spin of proton and neutron in this exchange process. Guided by analogy with the hydrogen molecule, Heisenberg assumed that the particles would "leave their spin directions behind", i.e. that after the exchange the spin of the neutron would point in the direction which was previously the spin direction of the proton and *vice versa*.

(3*c*) Majorana[5] proposed a law of force which was exactly the same as that of Heisenberg with the exception that he assumed the particles to retain their spins, i.e. after the exchange the proton, although it had now gone to the point P', would still have its spin pointing in the same direction as before.

Both the hypotheses of Heisenberg and of Majorana require, in order to be complete, a specification of the dependence of the exchange frequency on the distance between the two particles. The simplest assumption is again, as in the case of Wigner, 3*a*, that the exchange happens with appreciable frequency only if the distance between the particles is smaller than a certain length a, and that the frequency of the exchange reaches a constant value B/h for small distances.

Provided assumption (3*b*) or (3*c*) is accepted and the exchange frequency is known as a function of the distance, the wave equation describing the interaction between neutron and proton is uniquely defined and everything else follows from straightforward application of quantum mechanics without further assumption. The result is that the exchange energy will not always have the same sign, just as the valency forces between atoms may sometimes represent an attraction (like those between two hydrogen atoms forming a H_2 molecule) or a repulsion (like, e.g., those between two helium atoms). One finds that with Majorana's assumption (3*c*), the forces will produce attraction if the orbital angular momentum of the two particles about their common centre of gravity is even (in units of $h/2\pi$) and repulsion if it is odd. With Heisenberg's assumption the result depends on the relative spin directions of the particles. If the spins are parallel, the result is the same as above (if the spins point in the same direction it is obvious that the two assumptions must be equivalent), while for opposite spins the attraction and repulsion are reversed. The absolute magnitude of the interaction energy is equal to h times the exchange frequency, i.e. it vanishes for distances greater than a and becomes equal to B for small distances.

§ 3. BINDING ENERGY OF DEUTERON

If a neutron and proton are bound together to form a ${}^2_1\mathrm{H}$ nucleus* or deuteron in its normal state all three assumptions discussed above amount to the same. For in all three theories the ground state of the deuteron has no orbital angular momentum (only Majorana's assumption could allow a ${}^2_1\mathrm{H}$ nucleus with orbital momentum 1 by reversing the sign of the exchange function, which however does not seem plausible), and since experiments have shown that it has a total angular momentum of 1 unit, the spins of the two particles must be parallel. Then all three theories lead to attraction.

Moreover, we know from experiments the binding energy of the deuteron, because we know the energy of γ-rays necessary to disintegrate it[6]. This energy is roughly 2 million electron-volts. From this, however, we cannot conclude that the constant B, the maximum value of the interaction energy, must be about 2 million electron-volts. For according to quantum mechanics there will also be a kinetic energy of the two particles even in the state of lowest energy. This "zero-point energy" is familiar from the quantum theory of the harmonic oscillator and is closely connected with the uncertainty principle. For if a particle of mass M is forced to move in a limited region of linear dimension d its co-ordinate is known within an error of d and consequently the uncertainty principle requires that its

* We use left-hand subscripts and superscripts for atomic number and mass number.

momentum must fluctuate at least by $h/4\pi d$. Thus its average kinetic energy, which is equal to $1/2M$ times the mean square of the momentum, will be $h^2/(4\pi d)^2 2M$. In the case of the deuteron we have to insert $2a$ for the size and $M_H/2$ for the mass, since our consideration applies to the relative motion of the two particles, for which the reduced mass is $M_H/2$. Thus the kinetic energy is $h^2/(8\pi a)^2 M_H$ and the total energy becomes roughly[3]

$$E_0 = -B + h^2/(8\pi a)^2 M_H \qquad \text{.....(1).}$$

B can be determined from (1) if a is known. B will be the greater the shorter the range a of the forces. If, in particular, a is very small, B will be large compared with 2 million electron-volts, and will practically be equal to

$$B \sim h^2/(8\pi a)^2 M_H .$$

Now from the binding energy of ${}^3_1\mathrm{H}$ we cannot obtain a and B separately, nor can we distinguish between the assumptions 3*a*, *b*, *c*. More detailed information can be obtained from the scattering of neutrons by protons which depends upon the same law of force.

§ 4. SCATTERING OF NEUTRONS BY HYDROGEN

This process must also be discussed with the aid of quantum mechanics, for even for very fast neutrons of 20 million electron-volts the de Broglie wave-length is only $0{\cdot}8.10^{-12}$ cm. (approx.), i.e. comparable with nuclear dimensions, and diffraction must be of importance.

Two limiting cases give very simple results: If the neutrons are slow, i.e. if their wave-length is smaller than the range a of the forces, only those neutrons can be scattered which have no angular momentum about the centre of gravity, i.e. whose classical orbit would pass the neutron at a distance smaller than the de Broglie wave-length. For such "head on" collisions the first and last of the above assumptions (Wigner and Majorana) give the same result, viz. the scattering should be spherically symmetric for an observer moving with the centre of gravity[7]. The absolute magnitude of the scattering cross-section (defined as the total number of neutrons scattered by one proton from an incident beam carrying 1 neutron per sq. cm.) should be equal to

$$\frac{h^2}{\pi M_H (E_0 + \frac{1}{2}E)} \sim \frac{9.10^{-30}\ \text{erg. cm}^2}{E_0 + \frac{1}{2}E} \qquad \text{.....(2),}$$

where E is the neutron energy and E is the binding energy of ${}^3_1\mathrm{H}$, i.e. $E_0 \sim 2$ million electron-volts. Thus the cross-section should be constant for energies below about a million electron-volts. Heisenberg's assumption would lead to the same cross-section for those neutrons which happen to have their spins parallel to that of the proton with which they collide; but it would give different results for those striking a proton with opposite spin. Although, according to Heisenberg, there would be repulsion with opposite spin, the cross-section for this case could be written in the form (2), the constant E_0 having a different physical meaning. As the *a priori*

probability for two particles to have their spins opposite to each other is 1/4, the total cross-section will be, according to Heisenberg,

$$\frac{\hbar^2}{\pi M_H}\left\{\frac{3}{4}\frac{1}{(E_0+\frac{1}{2}E)}+\frac{1}{4}\frac{1}{(E_0'+\frac{1}{2}E)}\right\} \qquad \ldots\ldots(2').$$

This reduces to (2) if $E_0=E_0'$. Experiments on the scattering of slow neutrons by hydrogen show that their scattering cross-section increases about ten times between energies of a million volts and very small energies, in contradiction with (2). The formula (2′) could explain the present experimental data with a value of E_0' of about 30,000 volts. Such a very low value could never be obtained from Heisenberg's assumption (3*b*), for according to this assumption opposite spins should have a repulsion exactly equal in magnitude to the attraction of particles with parallel spin; therefore E_0' should be of the same order of magnitude as E_0, although numerically different from it. One can, however, obtain arbitrary values of E_0' by combining a force of the Majorana type with one of the Heisenberg type.* For these forces will work in the same direction for parallel spins, but in opposite directions for opposite spins. Therefore E_0 will be produced by their sum, while E_0' will be due to their difference. At present such an assumption seems to give the best description of the experimental facts. It is, however, also possible that the low value of E_0' is due to the magnetic interaction between the spins of the particles, which has so far been neglected.

The result that slow neutrons should show a spherically symmetric scattering relative to the centre of gravity seems to be in accord with experiments.

If the neutrons are very fast, i.e. if their wave-length is smaller than the range of the forces, the ordinary potential (Wigner) would produce only small deflections from their initial direction. The protons should then, according to the laws of conservation of momentum and energy, be knocked off with a small energy and almost at right angles to the neutron beam. With an exchange force, however (Heisenberg or Majorana), the particles would exchange their rôles, i.e. the neutron should be almost completely stopped and go off at right angles, while the proton would go on with the original neutron beam[7].

Such high energies are not practically available, but a beginning of this effect should be observed as soon as the wave-length becomes comparable with the range *a* of the interaction force. Thus an observation of the energies at which the scattering ceases to be isotropic would give an estimate of *a*, the sign of the anisotropy (i.e. whether a greater number of protons or neutrons go forward) would decide between the assumption (3*a*) on the one and (3*b*, *c*) on the other hand. So far such scattering experiments have only been done with limited accuracy and do not determine the sign of the anisotropy. The fact that no anisotropy could be found at a few million volts indicates that *a* is certainly smaller than 10^{-12} cm.

* This explanation was given in an unpublished remark by Wigner. Cf. also reference (7*a*).

§ 5. STRUCTURE OF DEUTERON

A very small value of a does not, however, imply that the mean distance of the two particles in a deuteron is very small. At distances greater than a the potential energy will vanish, i.e. it will be higher than the negative energy $-E_0$. In classical mechanics, therefore, the motion would be confined to distances smaller than a. In wave mechanics, however, configurations with an energy less than the potential energy are not absolutely disallowed. It is known that in such cases the de Broglie wave, as it enters the region of "negative kinetic energy", undergoes total reflection. The intensity of the wave (square of the amplitude) then drops to zero as $e^{-\alpha r}/r^2$ where r is the distance and the constant α has the value $\frac{2\pi}{h}\sqrt{(ME)}$, i.e. $\sim 2{\cdot}3 \cdot 10^{12}\,\text{cm.}^{-1}$

Thus, as a result of the wave nature of the particles, the mean distance will always be at least $0{\cdot}4 \cdot 10^{-12}$ cm., however small the range of the force. If the range is small the greatest part of the wave will lie outside the range of the force and its properties are therefore independent of a and of the particular form of the interaction function. Under this assumption, one can calculate the probability of the disintegration of ${}^3_1\text{H}$ by γ-rays (nuclear photoelectric effect) and the result[7,8] agrees with the observed value[6].

§ 6. FORCES BETWEEN EQUAL PARTICLES

For all other nuclei than ${}^3_1\text{H}$ the other forces, i.e. the proton-proton and neutron-neutron interaction, also come into play. According to Heisenberg[4], one can obtain some information about them from the regularities in heavier nuclei. For it is known that the atomic weight of nuclei is always approximately twice the atomic number, i.e. they contain an equal number of protons and neutrons. This means that the neutron-neutron force and the proton-proton force must be about equal, because if, say, the proton-proton interaction should lead to a much higher interaction energy (stronger repulsion or smaller attraction) than the neutron-neutron force, the stable nuclei would contain more than half neutrons and less protons. Some such effect is actually found, because the atomic weight is actually slightly more than twice the atomic number. This may be due to the electrostatic repulsion between the protons.

Apart from this small effect, one has to consider the forces to be symmetric. Moreover, the above-mentioned fact also excludes a very strong attraction between equal particles. If the attraction would be stronger than that between neutron and proton, stable nuclei would consist of either protons or neutrons only. Thus all one can assume for these forces is a weak attraction or a repulsion.

If we now turn to the other light nuclei, ${}^2_1\text{H}$ and ${}^3_2\text{He}$, consisting of a proton and two neutrons or a neutron and two protons respectively, we notice that their binding energies[2] (9 and 8 million electron-volts) are almost equal. The difference is probably within experimental error. This confirms again that the neutron-neutron

and the proton-proton forces are almost equal. Further we see that the binding energies of both these nuclei are considerably greater than that of ${}^{3}_{1}$H. This seems at first sight difficult to explain, as we have just seen that the interaction between equal particles, if it is an attraction, must be weaker than that between neutron and proton, while these nuclei contain only two neutron-proton bonds each. However, Wigner[3] has shown that this discrepancy can be removed even without any force acting between equal particles, from the fact that for small a the binding energy of the deuteron results as a small difference between kinetic and potential energy. For the ${}^{3}_{1}$H, the total potential energy should be twice as large as for the deuteron, since the proton can interact with two neutrons, while the kinetic energy, instead of being that of two particles should now be that of three particles compressed to about the same volume, and should be therefore roughly 1·5 times the kinetic energy of the deuteron. The compensation will therefore be less complete than in ${}^{3}_{1}$H and consequently the influence of the large attractive potential will lead to a much higher binding energy. The binding energy of the α-particle will be much larger still.

Wigner has shown that one can account for the mass defects of all these light nuclei by putting $a = 10^{-13}$ cm. This value is not in contradiction with the upper limit deduced from scattering experiments. The corresponding value of B would be ~100 million electron-volts. Recently, Feenberg[9] has taken into account the interaction between equal particles and has estimated a to be $1{\cdot}3 \cdot 10^{-13}$ cm. (i.e. $B \sim 87$ million electron-volts)[10].

§ 7. SATURATION OF BONDS

However, according to this picture, one would expect the binding energy to increase at this same rate as the nuclei become heavier, while empirically for nuclei heavier than the α-particle the energy per particle is approximately a constant (this is equivalent to saying that the atomic weights of pure isotopes, expressed in terms of ${}^{4}_{2}$He = 4 or ${}^{16}_{8}$O = 16, are very nearly integers).

At this point the models of Heisenberg and Majorana differ considerably from that of Wigner (3*a*). For here the forces are of the same type as chemical valency forces, and just like these they show saturation. If two hydrogen atoms are bound together in a molecule, they will not attract a third hydrogen atom, or at least only by forces of different origin, and they will show a strong repulsion if the other atom comes too near. The same phenomenon of saturation occurs in the nucleus if we assume either of the models (3*b*) or (3*c*)[5]. Each neutron in the nucleus will be attracted by a limited number of protons only, and similarly each proton will be attracted by only a limited number of neutrons. This number is 1 in Heisenberg's theory (3*b*) and 2 in Majorana's (3*c*). Accordingly, the force acting on each neutron will become constant, independent of the nuclear size if the nucleus contains more than one or two protons respectively. The saturation should then, according to Heisenberg, start already at ${}^{3}_{1}$H, while according to Majorana ${}^{4}_{2}$He should be the first saturated structure. The experimental evidence rules out the first possibility

and would be compatible with Majorana's model. Actually we have seen that the combination of both is likely to be near the truth and although the consequences of this assumption have not yet been worked out in detail, they seem to be compatible with the trend of the mass curve for nuclei of medium atomic weight.

To calculate from quantum mechanics the actual values of the binding energies of heavy nuclei or their excited levels will only be possible with a more detailed knowledge of the interaction forces.

§ 8. TRANSMUTATIONS

It is much easier, however, to apply quantum mechanics to nuclear reactions, i.e. spontaneous and artificial disintegrations. For these processes can be considered in three stages[11]: (1) the impinging particle (in artificial disintegrations) must come very near to the nucleus; (2) the nucleus must undergo a rearrangement of its constituents; (3) another particle must be emitted. Amongst these steps the second is the least known and it will, of course, sensitively depend on the particular reaction and on the interaction forces between the nuclear constituents. However, it has, in general, the least influence on the probability of a reaction. For the interaction energies between particles in a nucleus are of the same order of magnitude as the binding energies of these particles. Once a particle has succeeded in entering the nucleus it will cause such a thorough perturbation of its motion that it has an appreciable chance of producing any change that is compatible with conservation of energy, angular momentum and other "selection rules".

The first and third step, or at least one of them, will, in most cases, only happen with small probability, for two reasons: One reason is the small geometric size of the nucleus which requires the impinging particle to go a long way through matter before it once hits a nucleus; the other reason is the repulsive force between the electric charge of the nucleus and that of the impinging particle. The second reason does not apply to neutrons or to the case of bombardment with, or emission of, γ-rays. Therefore with heavy nuclei capture of neutrons with emission of γ-rays or *vice versa* are the only reactions which can be produced artificially.

§ 9. EMISSION OR CAPTURE OF CHARGED PARTICLES

The probability for a charged particle to cross the strong repulsive field when entering or leaving the nucleus can be calculated without knowing the behaviour of the nucleus in any detail, since we have seen that the nuclear forces have a very short range and therefore outside the nucleus the particle is subject to the Coulomb repulsion only.

In classical mechanics the particle could pass through this field only if its energy would be higher than the maximum value of the potential, i.e. the Coulomb potential at that distance at which the nuclear attraction comes into play. If r_0 denotes the nuclear radius, this maximum value is $ZZ'e^2/r_0$ (Z atomic number of the nucleus, Z' that of the particle, i.e. $Z'=1$ for a proton, and 2 for an α-particle, e = electronic charge). For heavy elements ($Z=90$, $r_0 \sim 10^{-12}$ cm.) this becomes

equal to about 15 million electron-volts. For not too fast particles and not too light nuclei, we therefore always have conditions in which classical mechanics would not allow the particle to reach the nucleus at all or, if it is in the nucleus, to get out.

The solution of this apparent difficulty by Gamow and by Gurney and Condon in 1928 represents the first application of quantum mechanics to nuclear problems: If the energy of the particle is below the maximum of the potential, the wave representing the particle will undergo total reflection, i.e. it will show an exponential decrease and the probability of finding the particle close to the nucleus will become small, but it will not vanish. The solution of the wave equation shows that the transmission coefficient for such a potential field (i.e. the probability of a particle reaching the nucleus, provided it arrives "head on") is, apart from a numerical factor which depends on the attractive potential near the nucleus, given by the expression

$$e^{-J} \qquad \text{.....(3)}$$

where J is the integral

$$J=\frac{4\pi}{h}\sqrt{(2M)}\int\sqrt{\{V(r)-E\}}\,dr \qquad \text{.....(3')}$$

to be taken over the region where $E<V$. Here M is the mass of the particle, $V(r)$ the potential energy, E the energy of the particle.

This expression governs, with certain exceptions, the probabilities of all nuclear reactions in which charged particles are emitted or absorbed* unless the particles are very fast or the nuclei very light.

One of the exceptions applies to those nuclei which, because of conservation of momentum, can absorb or emit only a particle with an angular momentum different from zero. For then the conservation of momentum forces the particle to spend a certain part of its energy on the angular motion and correspondingly less is left for the radial motion, i.e. for the crossing of the potential field. If $l.h/2\pi$ is the angular momentum, the transmission coefficient is the same as if the potential at each point at distance r were higher by $l(l+1)\,h^2/8\pi^2Mr^2$. In consequence of this the probability of such reaction will be considerably smaller than that of reactions in which only charged particles with no angular momentum are emitted[12].

The other exception refers to cases of resonance. If the potential barrier had no transmissibility at all, the particle would have some stationary states (quantized orbits) inside the nucleus, even if its energy is positive. For a finite but small transmission coefficient these states still persist in the sense that a particle with such an energy will stay in the nucleus a very long time although ultimately escaping. The time it spends in the nucleus is greater the smaller the transmission coefficient. If the energy of the particle happens to coincide with one of these so-called "virtual energy levels" it will have a greater chance of producing a nuclear reaction in spite of the small transmission coefficient, for although it has a small chance of entering the nucleus it will remain inside for a long time when once it gets there. The position of these virtual levels cannot, of course, be predicted without knowing sufficient

* A more detailed discussion of (4) and its comparison with observed data can be found in last year's report by Andrade[1]. Cf. also reference (11).

details of the nuclear forces to calculate the energy levels of nuclei, but their existence and the behaviour of the excitation function near such a level are found in agreement with quantum mechanics.

§ 10. CAPTURE AND SCATTERING OF NEUTRONS

In reactions in which the bombarding particle is a neutron and in which no charged particle of low energy has to be emitted, the governing factor for the reaction-yield is the probability that the neutron hits the nucleus. At first sight it might seem that this probability should be determined by the geometric cross-section of the nucleus since it is probable that all forces acting on a neutron have a short range. Then the apparent cross-section for any process involving capture or scattering of a neutron could at most be of the order of a nuclear cross-section, i.e. of the order 10^{-24} cm.^{2} Actually cross-sections as large as 10^{-20} have been observed. The explanation of this is again supplied by the wave properties of the neutrons, which may increase the cross-section by a phenomenon of resonance similar to the resonance for charged particles discussed above[13, 14].

A special case of this type has occurred already in formula (2) where, in spite of the short range of the interaction force, the cross-section for slow neutrons would become very large if E_0 should happen to be small. This may be interpreted in the following way: If E_0 vanished exactly, and if the neutrons had no kinetic energy outside the nucleus, they would just fit into a quantized orbit. Their density distribution would then have a sharp maximum near the nucleus and vanish at infinite distance, i.e. the neutrons would spend a long time very near the nucleus and consequently have a greater chance of reacting with it. If E_0 is small, but finite, and if the neutrons have a small finite energy, these effects become less pronounced, but the time the neutron spends near the nucleus will still be considerably larger than the time it spends in any other region of similar size. This fact offers a qualitative explanation of the large cross-sections observed for extremely slow neutrons. The detailed theory of the various reactions has yet to be carried through.

§ 11. β-EMISSION; THE NEUTRINO

So far we have considered only reactions in which only heavy particles take part. One knows, however, that nuclei can in certain cases emit electrons (β-decay) or positive electrons. Where do they come from? If we retain the view that all nuclei consist of neutrons and protons it is obvious that after the β-decay the new nucleus, since it has a higher charge, but the same mass, must contain one proton more and one neutron less than the original nucleus. In other words a neutron must have been transformed into a proton with emission of an electron.

This view, however, meets with serious difficulties[1]. The most important one is that both electron and proton have a spin of $\frac{1}{2}$ unit (i.e. $\frac{1}{2}h/2\pi$). Since their orbital angular momentum is always an integral number of units, the total spin must also be an integer. On the other hand the angular momenta of all nuclei require half integral values for the spin of the neutron. We would have to conclude

therefore that the process of β-decay violates the principle of conservation of angular momentum. Similarly the process seems to violate the conservation of energy. For the electrons emitted by the same radio-active element, instead of all having the same energy, cover a wide range of energies continuously. But the initial and the final nucleus are in perfectly definite states and no difference between the physical states of different nuclei of the same element can be observed. Therefore one must conclude that the energy difference between the initial and the final nucleus is a definite quantity, and consequently is not always equal to the energy of the electron which is different for different individual electrons. Thus the total energy does not seem to be conserved in the process.

While the question of whether one actually must abandon the conservation laws is not settled yet, it seems probable that the right explanation is that proposed by Pauli[15], according to whom a new particle of mass smaller than, or equal to, the electron mass and with no charge, is emitted simultaneously with the electron. This particle would then take away some kinetic energy and thus make up the energy balance; if one assumes it to have a half integer spin (probably half a unit) the angular momentum also would balance in the process. One can show[16] that this particle, the so-called "neutrino", would have a very small action indeed on other nuclei after it has been emitted and a direct proof of its existence seems very difficult.

The most natural view would then be that the neutron consists of a proton, an electron and a neutrino. On the other hand, however, we know that artificial radio-active elements can disintegrate with the emission of a positive electron (and probably a neutrino) which would rather indicate that the proton should consist of a neutron, a positive electron and a neutrino. Since both proton and neutron cannot at the same time consist of each other, and since between the emission of positive and negative electrons there seems to be complete symmetry, one is led to the idea that neither of them is a complex structure containing electrons[17]. Fermi[18] proposed that both neutron and proton are elementary particles and that the neutrino and the electron are created in the process of β-decay, just as a light quantum is created in the process of light emission by an excited atom. A quantitative theory suggested by Fermi[18] to describe such a phenomenon seems to be in reasonable agreement with both the form of the energy distribution curve for the emitted β-electrons and with the connexion between the periods of β-emitting nuclei and their maximum β-energies.

It has further been suggested[20,21] that the forces leading to the β-decay are also responsible for the attraction between neutron and proton, in the same way as the electromagnetic field created by an atom accounts equally for the light emission and for the attraction between different atoms. We have seen, indeed, that the interaction between proton and neutron contains an exchange process in which the particles change rôles. This might then be connected with the emission of an electron and a neutrino by the neutron (which thus becomes a proton) and re-absorption of these particles by the proton, turning the latter into a neutron. To reduce in this way two different phenomena to a common origin would be very satisfactory. There are still, however, quantitative difficulties, as the two phenomena

seem to require a different order of magnitude for the forces leading to creation and absorption of electrons. Since, however, very little is known about the nature of these forces, and since the relative order of magnitude of β-emission and proton-neutron interaction is very sensitive to the assumptions made regarding the velocity dependence of these forces[22], the difficulties may still be overcome.

Part 2. THE SPECTRA OF POLYATOMIC MOLECULES

By G. B. B. M. Sutherland, Pembroke College, Cambridge

§ 1. INTRODUCTION

Since the interpretation of the hydrogen spectrum by Bohr in 1913, it has generally been taken for granted that it will only be a matter of time until any spectrum we care to produce can be analysed to give the energy levels of the system emitting (or absorbing) that spectrum. Subsequent developments are justifying that expectation at a quite extraordinary speed. Within the short space of twenty years the spectra of practically all the elements have been elucidated, yielding information on the electronic configurations of the atoms, which is an invaluable asset in a host of physical problems. The spectra of over 150 diatomic molecules have already been analysed, but hitherto the progress made with the spectra of polyatomic molecules has been disappointingly slow. Indeed, up to the present there is no polyatomic molecule of which we know the energy levels with any degree of completeness, although the spectra of several have been partially analysed. The knowledge derived from such partial analyses is nevertheless of considerable value and since it may well be that we shall have to be content with this for some time to come, it is important to know the exact nature of the factors which are militating against a more complete knowledge. Although the slow rate of progress in polyatomic spectra has been partly due to these factors, it has also been due to the natural tendency of the investigator to move from the spectra of simple atomic systems to those of larger molecular systems, only after the former have been fairly well analysed. There is little doubt that the next ten years will bring about a tremendous advance in our knowledge of the spectra of polyatomic molecules. The purpose of this review is to show how the steps have been prepared for this advance, and to explain generally the foundations on which the analyses of polyatomic spectra will proceed.

It is well known that the energy of any diatomic molecule may be written as the sum of three practically independent terms, viz. the electronic energy E_e, the vibrational energy of the nuclei, E_v, and the rotational energy of the molecule as a whole, E_r, thus:

$$E = E_e + E_v + E_r \qquad \text{.....(1).}$$

Then if E' denotes the initial state of the molecule, and E'' the final state (after emission or absorption of radiation) the frequencies in the spectrum observed are given by

$$h\nu = E' - E'' = (E_e' - E_e'') + (E_v' - E_v'') + (E_r' - E_r'') \quad \ldots\ldots(2).$$

Since, in general, $E_e \gg E_v \gg E_r$, the first term determines the position of the band spectrum, the second is responsible for the banded structure, while the last gives rise to the fine structure within each separate band. It may happen that $E_e' = E_e''$, i.e. that there is no change in the electronic state of the molecule, in which case the second term becomes the controlling one as regards the position. This is known as a vibration-rotation spectrum and always falls in the infra-red region of the spectrum. If in addition $E_v' = E_v''$ then we obtain a pure rotation spectrum, which lies in the very far infra-red. In diatomic spectra the two latter types are seldom investigated, since the information they yield can usually be more accurately obtained from the electronic band spectrum.

All of this general picture applies equally well to the polyatomic molecule. The particular difficulties occur when we come to consider the explicit expressions for the different energy terms E_e, E_v and E_r and the rules governing the changes in the quantum numbers characterizing these energies. For instance, in a diatomic molecule there is only one vibrational quantum number (v) associated with E_v and Δv may equal any positive integer, but for a polyatomic molecule there will be several quantum numbers characterizing E_v, and the selection rules associated with them can all be different; similar complexities occur for the rotational energy. The electronic spectrum of any polyatomic molecule is therefore so complicated that it is more profitable to start with the simpler vibration-rotation and pure rotation spectra, and we shall follow this method in our treatment here, beginning with the simplest kind of spectrum—the pure rotation spectrum—and proceeding to the discussion of vibration-rotation spectra and, finally, of electronic spectra.

Before starting the theoretical discussion of these three types of spectra it will be advisable to say a little about their production. Pure rotation spectra and vibration-rotation spectra are obtained either as absorption spectra (in the infra-red) or as Raman spectra, since emission spectra of such a kind are extremely difficult to obtain experimentally, and are unlikely to yield information which is unobtainable from absorption or Raman spectra. It is important to realize the exact difference between an absorption spectrum and a Raman spectrum. In the former case, the molecule absorbs energy when the frequency of the incident radiation coincides with the energy difference between two of its stationary states, and such a transition is associated with a change in the permanent electric moment of the molecule. In the latter case, the molecule scatters an optical frequency bearing no relation to its energy levels at all; in the scattering process, the incident frequency may be unaltered (classical Rayleigh scattering), may be diminished, or may be enhanced. The Raman spectrum is given by the frequency differences between the altered frequencies and the incident frequency; these are found to correspond to vibration and rotation frequencies of the scattering molecule. In this case, therefore, the transition of the molecule between two of its stationary states is associated with a

change in its induced moment because the intensity of scattering depends on the moment induced in the molecule. Since, moreover, the moment induced depends on the polarizability of the molecule, the appearance of any particular frequency in the Raman spectrum of a molecule will be determined by the way in which the polarizability is affected by that particular frequency. It is clear, then, that there will be fundamental differences between the absorption spectrum and the Raman spectrum of any molecule. Although the same frequencies may occasionally appear in each, this will be the exception rather than the rule. It was hoped, at first, that the discovery of the Raman effect would make it unnecessary to investigate the infra-red spectra of molecules, but this has proved illusory and a knowledge of both is required in order to determine the complete system of vibrational and rotational energy levels of the majority of molecules.

Electronic spectra are generally observed in absorption, since any method of exciting such spectra intensely in emission is usually so powerful that it dissociates the molecule. Unfortunately the electronic spectra of many of the simpler polyatomic molecules lie in the very far ultra-violet where high dispersion is difficult, but recent advances in the technique of this region[1] are already overcoming this difficulty.

§ 2. PURE ROTATION SPECTRA

The problem of predicting the rotational energy states of any molecule and the transitions between them which will be active in infra-red and Raman spectra may be considered as completely solved. The only data required are the moments of inertia of the molecule, and so conversely we expect to be able to deduce the moments of inertia of any molecule from an analysis of its rotation spectrum. It is found that the rotational energy may usually be expressed as a function of the moments of inertia and of one, or two, rotational quantum numbers, and that the transitions between the separate energy states are governed by selection rules between these quantum numbers. The energy expressions and the associated selection rules become progressively more complicated as we pass from the case of the linear molecule (with one effective moment of inertia) to that of the asymmetrical top molecule (with three different moments of inertia). We have here, in fact, a natural and convenient way of classifying polyatomic molecules according to their degree of rotational symmetry. This is illustrated by table 1 which gives the energy formulae applicable in the four possible cases, together with the appropriate selection rules for transitions which are active (*a*) in infra-red absorption, and (*b*) in Raman scattering.

Thus for a linear molecule the frequencies of the lines in its infra-red absorption spectrum will be given by

$$h\nu_{\text{rot}} = E_r' - E_r'' = \frac{h^2}{4\pi^2 A} j \qquad (j = 1, 2, 3, \ldots) \qquad \ldots\ldots(3).$$

In other words, the spectrum is exactly like that of a diatomic molecule, consisting of a series of equally spaced lines, having a frequency difference between them inversely proportional to the moment of inertia of the molecule. The position of such

a spectrum in the infra-red clearly depends on the magnitude of the moment of inertia, lying farther out the greater the moment. For instance, a molecule like N_2O will have its rotation spectrum stretching roughly from 200 to 2000μ., and the separation between the lines is 0·8 cm^{-1}. Now although infra-red measurements have been made as far as 250μ. with fairly high dispersion, the technique is very difficult and it is therefore not surprising that so far no linear polyatomic molecule has been investigated by this method. On the other hand, the rotational Raman spectrum, although not very easy to obtain (since the molecule must be in the gaseous state), will give just as much information, and with greater accuracy since

Table 1. Rotational-energy formulae and selection rules for the four classes of polyatomic molecules

Class of molecule (A, B, C denote the moments of inertia of the molecule)	Expression for the rotational energy E_{rot}	Selection rules	
		(a) Infra-red spectra	(b) Raman spectra
Linear $A=B$; $C=0$	$\frac{h^2}{8\pi^2 A} j(j+1)$ $j=0, 1, 2, \ldots$	$\Delta j = \pm 1$	$\Delta j = 0, \pm 2$
Spherical $A=B=C$	$\frac{h^2}{8\pi^2 A} j(j+1)$ $j=0, 1, 2, \ldots$	No active transitions	No active transitions
Symmetrical top $A=B\neq C$	$\frac{h^2}{8\pi^2 A} j(j+1) + \left(\frac{1}{C}-\frac{1}{A}\right)\frac{h^2 k^2}{8\pi^2}$ $\lvert k \rvert \leqslant j; j=0, 1, \ldots$	$\Delta j = \pm 1$ $\Delta k = 0$	$\Delta j = 0, \pm 1, \pm 2$ $\Delta k = 0$
Asymmetrical top $A\neq B\neq C$	No simple explicit expression in terms of j, the quantum number for total angular momentum	$\Delta j = \pm 1$	$\Delta j = 0, \pm 1, \pm 2$

(a) the dispersion available is greater in the optical region, and (b) the spacing between the lines is double ($\Delta j = 2$) that for the infra-red spectrum ($\Delta j = 1$). Again, only a molecule which has a permanent dipole moment will give rise to a pure rotation spectrum in the infra-red, whereas every linear molecule possesses a rotational Raman spectrum. Examples of linear polyatomic molecules of which the Raman spectra have been successfully obtained are carbon dioxide (O—C—O) and acetylene (H—C—C—H).

It should be noticed that spherically symmetrical molecules (e.g. methane, CH_4, carbon tetrachloride, CCl_4, etc.) possess no pure rotation spectra.

For symmetrical top molecules, an application of the selection rules to the energy formula given in table 1 shows that in the infra-red the spectrum is no more complex than that of a diatomic molecule of moment of inertia A. Thus an investigation of the spacing in such a spectrum can yield no information about the moment of inertia (C) around the symmetry axis of the molecule. The same is true of the Raman spectrum in which the spacing of the lines is exactly the same as in

the infra-red spectrum. The molecules of this class which have been investigated up to date are those of ammonia, NH_3 (infra-red and Raman), of phosphine, PH_3, and of arsine, AsH_3 (infra-red). Although the spacing between the lines only gives one of the moments of inertia of the molecule (A), the intensities of the lines can be shown to depend on both the moments of inertia and consequently an accurate measurement of them allows at least a rough determination of the dimensions of the molecule to be made.

The case of the asymmetrical top molecule is more complicated. The energy depends on the total angular momentum of the molecule (characterized by the quantum number j) and on the three moments of inertia, but it is not possible to give an explicit expression for it, although the energy levels may all be computed by solving certain algebraic equations. Each value of j is associated with $2j+1$ levels and necessitates a separate computation. The question of the selection rules is also rather complicated, for in addition to the simple one given in table 1, there are additional rules governing the transitions between the subsidiary levels of each j value. It should be emphasized that the computations can only be done when the moments of inertia are known. This seems rather like begging the question, but in fact the approximate values of the moments of inertia can usually be deduced from other sources and so by gradually bridging the gap between calculated and observed values it is theoretically possible to arrive at accurate values for the moments. Such a procedure of successive approximation is rather laborious and requires very accurate experimental data. It has not been carried out accurately even for the water molecule, the pure rotation spectrum of which is much the best known of asymmetrical molecules. Mecke[2], however, has shown that all of the known lines of water vapour may be classified by treating the asymmetrical rotator as an imperfect symmetrical rotator.

It seems unlikely that pure rotation spectra are going to occupy an important place in the development of the spectra of polyatomic molecules. Although they are very simple in structure compared with vibrational and electronic spectra, the experimental difficulties in their detection are so much greater and the information they provide is relatively so much less that investigators will tend to neglect them. Moreover, the information obtainable from them is nearly always obtainable much more accurately from vibrational spectra. They will be useful, however, in cases where the vibration-rotation spectra are very complex, and interpretations are ambiguous, and where it is specially desirable to obtain a knowledge of the constants of the ground state of the molecule unaffected by vibration or other excitation of the molecule.

§ 3. VIBRATION-ROTATION SPECTRA

The vibrational energy levels of a diatomic molecule are given (to a first approximation) by those of a simple harmonic oscillator, viz.

$$E_v=(n+\tfrac{1}{2})\,h\nu \quad (n=0,\,1,\,2,\,\ldots) \qquad \ldots\ldots(4),$$

where

$$\nu=\frac{1}{2\pi}\sqrt{\frac{k\,(m_1+m_2)}{m_1 m_2}} \qquad \ldots\ldots(5).$$

In other words we know the vibrational energy as soon as we know the mechanical frequency of the oscillator formed by the two nuclei of masses m_1 and m_2 constrained in their equilibrium position by a restoring force given by k. In the same way we obtain the first approximation to the vibrational energy of a polyatomic molecule, by regarding it as a vibrating system of nuclei and determining its mechanical frequencies. Thus the general polyatomic molecule containing n atoms has $3n$ degrees of freedom, of which 6 are accounted for by the translational (3) and rotational (3) motions of the molecule as a whole, leaving $3n-6$ to be accounted for by internal or vibrational motions of the nuclei. The interpretation of this on classical mechanics is that such a system possesses $3n-6$ *fundamental* or *normal* modes of vibration from which every possible mode of vibration of the system may be built up. The values of the corresponding *normal frequencies* may be calculated provided the masses of the atoms are known and the constants of the force field which controls their vibrations. In quantum mechanics this means that the vibrational energy levels of the molecule are formed additively from those of the $3n-6$ harmonic oscillators which correspond to the normal vibrations. So we write

$$E_v = \sum_1^{3n-6} (n_1 + \tfrac{1}{2})\, h\nu_1 \qquad \text{.....(6),}$$

where each of the $3n-6$ frequencies ν may formally be expressed as

$$\nu_1 = \frac{1}{2\pi}\sqrt{\frac{f(k_1 k_2 \ldots)}{F(m_1 m_2 \ldots)}} \qquad \text{.....(7).}$$

It will occasionally happen that some of these $3n-6$ vibrations coalesce in twos or threes to form what are known as doubly, or triply, degenerate vibrations. This will be the case whenever the motions of the atoms in executing the vibrations are isotropic in a plane (doubly degenerate) or in space (triply degenerate). For instance, the vibrations ν_4 and ν_5 of the acetylene molecule (figure 1) are each "double" since the motions of the atoms are isotropic in planes perpendicular to the axis of the molecule. In this way a highly symmetrical molecule may have quite few normal modes of vibration, although it has a great many degress of vibrational freedom. The fact that methane, CH_4, is so symmetrical reduces the number of its fundamental frequencies to 4 although it has altogether 9 degrees of vibrational freedom. It follows that the spectrum of a highly symmetrical molecule will be simpler than that of an unsymmetrical one.

Just as in the diatomic case the actual energy levels are not given accurately by (4), but require correction for anharmonicity and interaction with rotation, so in the polyatomic case (6) only represents the first approximation to the true energy expression. A more accurate one would have the form:

$$\begin{aligned} E = {} & n_1\nu_1 + n_2\nu_2 + \ldots \\ & + n_1^2 x_{11} + n_2^2 x_{22} + \ldots \\ & + n_1 n_2 x_{12} + n_1 n_3 x_{13} + \ldots \end{aligned} \qquad \text{.....(8),}$$

where terms of the form $n_1 x^2_{11}$ take account of the pure anharmonicity of each vibration while those of the form $n_1 n_2 x_{12}$ are due to interaction between the various vibrations. The problem of the interaction of rotation and vibration in polyatomic

molecules is a rather difficult one, and has been solved only for a few special cases[3]. For present purposes it will be sufficient to assume that the energies of vibration and rotation are additive and independent and so the energy of a vibrating rotating molecule is given by the sum of an expression of the form (8) and the appropriate formula from table 1. The vibration-rotation spectra (infra-red and Raman) are due to transitions between these levels. At first sight it might seem an almost hopeless task to elucidate such spectra (unless for very simple molecules), but there are many simplifying circumstances which render the analysis much easier than would be anticipated. These simplifications are due, partly to the intrinsic characteristics of such spectra, but mainly to the symmetry properties of the normal vibrations.

Thus, the analysis of Raman spectra is greatly facilitated by the knowledge that only transitions from the ground state to the *first* of any of the excited vibrational levels are to be expected in Raman scattering; that is, only fundamental frequencies are observed and no overtone or combination frequencies. Although this rule is fairly well established experimentally it is not a strict selection rule. All that is implied is that the intensities of overtone and combination frequencies (such as $2\nu_1$, $\nu_1+3\nu_2$) are of such a different order of magnitude from those of the fundamentals as to be unobservable in practice. As a matter of fact, in certain anomalous cases, where an overtone frequency "resonates" with an active fundamental its intensity may become comparable with that of the fundamental. In general, however, it is reasonable to conclude that any frequency which is observed in a Raman spectrum is a fundamental vibration frequency of the scattering molecule. No similar rule applies to infra-red spectra. Here overtone and combination frequencies are always active, and the intensity of some of them may be comparable to that of an active fundamental, so that it is not always safe to conclude that the most intense absorption bands correspond to the fundamental frequencies of the molecule.

One of the most helpful principles in the analysis of vibration spectra is the purely empirical one of decomposing the vibrations of the molecule into quasi-constituent vibrations of smaller groups in the molecule. This means that although, strictly speaking, each fundamental frequency is concerned with a vibration of the molecule as a whole, in actual fact it is often confined principally to a vibration within one group of the molecule, or to a vibration of one group in the molecule with respect to another. For instance, the CH_2 group has three fundamental frequencies and whenever we find that group in a larger molecule we shall expect to find those three characteristic frequencies modified only by a few wave numbers[4]. In fact the normal vibrations of any molecule may always be approximated by this method of building them up from simpler units in the molecule. It should be clearly understood, however, that this is only a partial description of the true vibration. The latter must be derived by a proper dynamical investigation of the complete vibrating system by the classical method of small vibrations. The importance of the former crude physical method is that it does give correctly the symmetry properties of the normal vibrations, and for establishing selection rules, it is the symmetry properties which are the really significant characteristics of a vibration.

Since the selection rules for vibration frequencies depend on symmetry properties of the vibrations it is almost impossible to give any general rules; one must consider particular molecules or particular classes of molecules. Here we shall limit ourselves to one molecule to illustrate the method. Let us consider the linear symmetrical molecule $XYYX$, such as acetylene (HCCH), and look on it as built up of two YX groups. This molecule will have seven $(3\times4-5)$ fundamental frequencies. The first (ν_1) may be described as the oscillation of the two groups with respect to each other along the line of the molecule (figure 1). The second and third may be built up from the vibrations of the two CH groups, taken respectively in phase (ν_2), and out of phase (ν_3). The remaining four must be accounted for by motions of the atoms perpendicular to the axis of the molecule. It is readily seen that they are given by the relative vibrations of the CH groups perpendicular to the line of the molecule, again taken in phase (ν_4), and out of phase (ν_5). The last two vibrations are doubly degenerate since the motions of the atoms are isotropic in planes perpendicular to the axis of the molecule. The way in which the vibrations ν_2, ν_3 and ν_4, ν_5 form pairs is quite a general feature of molecules having a centre of symmetry. It will be observed that one of each pair is "symmetrical" with respect to the centre of symmetry while the other is "antisymmetrical". "Symmetrical" here means that on reflection in the centre of symmetry the phase of the vibration is unaltered; "Antisymmetrical" means that on reflection the phase of the vibration is altered by π. For such a molecule the rule governing the appearance of fundamentals is very simple, viz. those vibrations which are symmetrical are active in the Raman spectrum but inactive in absorption, while the opposite is true for the antisymmetrical vibrations. This is a perfectly general rule (one of the few); it follows that if for any molecule, a frequency is observed both in the absorption and in the Raman spectrum then that molecule cannot possess a centre of symmetry. The rules for molecules having a lower degree of symmetry cannot be so concisely stated and would require a more detailed treatment of the various symmetry types than we have space for here. Sufficient to say that the rules for all possible types have been worked out by Dennison[5] and by Tisza[6] for the infra-red and by Placzek[7] for the Raman effect. These rules concern not only fundamental frequencies but also overtone and combination frequencies. The rules for the Raman effect give in addition the polarization properties of the various frequencies that are allowed. These are of considerable value in assigning fundamental frequencies to particular modes of vibration of the molecule in the cases where polarization measurements have been made. Unfortunately this is done very little at present but it is to be hoped that it will become more general in the future.

So far we have only considered the factors governing the appearance or non-appearance of particular frequencies in a vibration spectrum, without regard to the accompanying changes in rotational energy which give rise to a fine structure in every vibration "band". The importance of this fine structure is twofold; it is an invaluable help in enabling us to assign the observed frequency to a particular mode of vibration of the molecule, and its analysis leads to the values for the moments of inertia of the molecule. It is for such reasons that infra-red absorption spectra are

so much more valuable in general than Raman spectra, because the rotational fine structure is readily observable in the former but extremely hard to obtain for the latter. It is easy enough to obtain the position of the Raman band because that can be done in the liquid, but in order to get the rotational fine structure the substance must be in the gaseous state and this presents quite a difficult problem in Raman spectroscopy, on account of the weakness of the scattered light from a gas. On the other hand, infra-red absorption spectra are normally done with the substance in the gaseous state. Yet even here the limitations of dispersion are quite serious. For instance the rotational fine structure of such a small molecule as NNO with a moment of inertia of 66×10^{-40} gm.cm.^{2} is getting quite near the present limit of resolving power in the infra-red, and yet it is a linear molecule with the simplest structure of all. The complete description of the fine structure of the vibration bands of the various classes of molecules has been given by Dennison[5] and others[3]; here we can only indicate the main features of each type.

The selection rules for the rotational energy changes are found to depend on the character of the vibration with which they are associated. For a linear molecule such as acetylene, there is a distinction between the rules for a "parallel" vibration (for which $\Delta j = \pm 1$) and those for a "perpendicular" vibration (for which $\Delta j = \pm 1, 0$). A "parallel" vibration means one in which the change of electric moment takes place parallel to the axis of the molecule (such as ν_3 in figure 1), while a perpendicular vibration denotes one in which the change of moment takes place perpendicular to the axis of the molecule (such as ν_4 in figure 1). The only difference in appearance of the bands is that in the parallel bands there is no Q branch ($\Delta j = 0$) at the centre of the band; otherwise they are identical, consisting of simple P and R branches ($\Delta j = \pm 1$) symmetrically disposed about the centre of the band, the spacing between the individual rotation lines being $h/4\pi^2 A$ cm^{-1}. If in the case of very heavy molecules, the line spacing (being inversely proportional to A) cannot be resolved, it is still possible to get a rough estimate for the moment of inertia by measuring the separation between the maximum intensity of the P branch and that of the R branch, and using the classical expression for the "doublet separation" of a rotating oscillator, viz. $\pi^{-1}\sqrt{(kT/A)}$.*

There is a similar distinction between the "parallel" and "perpendicular" bands of a symmetrical top molecule, according as the vibration takes place along or perpendicular to the symmetry axis of the molecule. For the parallel bands the selection rules are $\Delta j = \pm 1, 0, \Delta k = 0$, giving a structure exactly the same as that of the perpendicular band of a linear molecule of moment of inertia A. The measurement of the line spacing in such a band can therefore yield only one of the moments of inertia (A) of the molecule. The selection rules for the perpendicular bands are $\Delta j = \pm 1, 0; \Delta k = \pm 1$ which lead to a number of overlapping bands, each of which may be looked on as due to a particular change in k associated with $\Delta j = \pm 1, 0$. The resulting structure is much too complicated to be resolved under present conditions. It is sometimes possible to pick out the Q branches ($\Delta j = 0; \Delta k = \pm 1$) of the subsidiary bands and from the spacing between them to calculate the other

* Here k is Boltzmann's constant and T is the absolute temperature.

moment of inertia C. In this last calculation careful account must be taken of the peculiar coupling which exists between the rotation and vibration in this type of molecule[3]. When the molecule is so large that the individual lines of the parallel bands cannot be resolved, it is still possible to obtain a fair estimate of the moments of the molecule from the spacing and the relative intensities of the maxima of the *P*, *Q*, and *R* branches by employing a modified form of the doublet separation expression used for linear molecules.

There being no axis of rotational symmetry in the asymmetrical molecule the terms "parallel" and "perpendicular" can no longer be applied. Instead we get three kinds of vibration band according as the change of electric moment takes place along the greatest, the middle, or the least axis of inertia. The selection rules and the resulting rotational structure are extremely complex, as we have already remarked in dealing with the pure rotation spectrum. Yet an analysis has now been

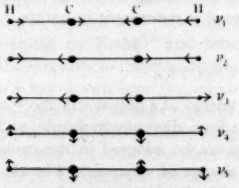


Figure 1. The normal vibrations of the acetylene molecule.

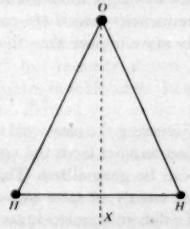


Figure 2. Symmetry of the water molecule.

given of the bands of H_2O and of H_2S, which yields the moments of inertia with great precision. For heavier molecules it has proved impossible to resolve the individual rotation lines and there is no approximate method by which the moments of inertia may even be roughly estimated by measuring the separation of intensity maxima. It would seem at present that there is a very definite limit to the number of asymmetrical top molecules that we can hope to tackle successfully in this manner, because of the excessive number of lines involved and the necessity of resolving them clearly before any information can be gained about the moments of inertia. Perhaps we shall have to depend on the analysis of the electronic spectra in the optical region,* or more probably on the results of X-ray and electron diffraction measurements in order to arrive at the dimensions of such molecules.

The main problem of the vibration spectra of polyatomic molecules may therefore also be considered formally as solved. Given the vibration spectrum of any molecule, it should be possible by an application of the selection rules to determine the symmetry class to which it belongs, its fundamental frequencies, and finally,

* It is significant that in the case of the two molecules H_2O and H_2S the overtone bands which fall in the photographic region were the key to the analysis.

(from an examination of the rotational fine structure) its form and moments of inertia. There is still one serious gap in our present knowledge of these spectra; that is in regard to the intensities of the allowed frequencies. It frequently happens that certain fundamentals which are "allowed" are not observed, or are so weak as to be apparently "forbidden". This is not yet properly understood, and so the above rules give only the maximum number of fundamentals which may be observed. The actual number will often be less. In the Raman effect this is sometimes expressed by the empirical rule that the highly symmetrical frequencies are the most intense. Thus in acetylene, ν_1 and ν_2 are found to be much more intense in the Raman spectrum than ν_5. This rule is explicable in a qualitative way when we remember that the intensity of a line in the Raman spectrum depends on the associated change in the polarizability of the molecule. It will usually be found that the highly symmetrical frequencies are more likely to cause a large change in the polarizability of the molecule than the less symmetrical ones. For a similar reason those frequencies which are confined mainly to stretching a homopolar bond are generally more intense than those where the bond is more ionic in nature.

§ 4. ELECTRONIC SPECTRA

In discussing the rotational and vibrational energy of a polyatomic molecule it is convenient to start from the energy expression for a diatomic molecule, and show how it can be generalized. This procedure cannot be adopted in the case of the electronic energy, at least without a complete change of viewpoint. The electronic states of a diatomic molecule are conventionally described by means of two quantum numbers Λ and S, the former giving the angular momentum of the electrons about the inter-nuclear axis, while the latter gives the resultant spin momentum of the electrons. Since, for the general polyatomic molecule, there is no preferred axis, with respect to which the electronic angular momentum is quantized, this method of description cannot be applied. Moreover, the way in which the electronic levels of polyatomic molecules are ordered cannot be given a simple physical interpretation. It depends on the symmetry properties of the wave functions associated with the various levels, which we can best understand by taking a concrete example, such as that of the water molecule (figure 2).

The form of this molecule is now well established from investigations on its vibration-rotation spectrum in the infra-red. It is known to be isosceles and consequently to have certain elements of symmetry. For example, rotation through π about the axis OX leaves it unchanged; so does reflection in the plane through OX and perpendicular to that of the molecule. If, then, ψ_1 be the wave function describing one of the electronic states of water, and we perform on its coordinates one of the above symmetry operations, it will be found to be modified in some simple way, such as to be multiplied by -1, or 2. Again, if ψ_2 describes another of the electronic states, it will be similarly modified on being subjected to such symmetry operations. It might be supposed that each of the wave functions would react in its own particular way to the operations of the symmetry group to which H_2O belongs, but

an application of group theory shows that actually there can only be four possible types of electronic wave function for this type of molecule. Thus, all the electronic states of water must belong to one or other of four classes—the wave functions of each class behaving in a characteristic way, when they are subjected to the operations of the symmetry group.

This is then the principle by which the electronic states of polyatomic molecules are classified. It is actually an extension of the way in which those of a diatomic molecule are described, when the latter are also treated by the methods of group theory. Thus all Σ states ($\Lambda = 0$) transform in the same way under a symmetry operation of the group to which the diatomic molecule belongs, and in quite a different way from Π states ($\Lambda = 1$) and so on.* The possible electronic states of any polyatomic molecule can therefore be found as soon as we know the symmetry group to which it belongs. In the language of the group theory, the number of different types of states is equal to the number of "irreducible representations" of the group. The irreducible representations of the 32 well-known symmetry groups occurring in crystals and their significance in classifying electronic levels were first demonstrated by Bethe[9] and others; Mulliken[10] has recently shown how to apply these results to the classification of electronic states in molecules. In addition Mulliken has shown how the selection rules may be derived very easily for transitions between the various types of level. Thus for the symmetry type C_{4v}† there are five possible kinds of levels designated by A_1, A_2, B_1, B_2 and E, but transitions from B_1 states can only take place to other B_1 states or to E states.

It should be observed that the form of the molecule must be known before its electronic states can be described. The form of the molecule may be deduced from its vibration spectra or (less certainly) from a detailed consideration of its possible electronic structure.‡ The real difficulty arises when we come to the ordering of these possible types of states in an energy table. The formal application of the group theory gives no information on this point, which requires to be carefully elucidated by a semi-empirical correlation with the diatomic case, and by examining the actual spectra in the light of the selection rules[11]. By such means Mulliken[11] has been able to identify the ground state and the first excited state of ethylene, while Eastwood and Snow[12] have extended the method to the treatment of formaldehyde and the aldehydes in general. In this way, it is possible to tell fairly accurately in which part of the molecule excitation is taking place when energy is absorbed—a matter of great importance in the study of photochemical reactions. Formerly such questions were solved in a qualitative experimental way by examining the spectra of a series of similar compounds. Thus, for the various aldehydes, an absorption is always observed in the same region of the spectrum (λ 2900 A.) and with roughly the same vibrational structure. It was generally assumed that this was connected with the excitation of an electron in the C—O bond which is common

* The correlation of the diatomic and the polyatomic molecule in this treatment has been very clearly explained by Lennard-Jones[8].

† This denotes a molecule with one axis of fourfold symmetry and four planes of symmetry through that axis.

‡ See section on Valency by Dr W. G. Penney.

to all aldehydes. Eastwood and Snow however have shown that the excitation is more likely to be correctly associated with one of the carbon electrons on the side removed from the O atom. In order, therefore, to get a complete understanding of the spectrum the formal description of the states and their associated selection rules must be accompanied by a careful consideration of the electronic structure of the molecule by the "orbital method" of Mulliken.

The vibrational structure of the electronic bands is only now beginning to be properly understood. The selection rules applicable to vibration-rotation spectra are no longer valid here, but Herzberg and Teller[13] have investigated the rules governing vibrational transitions in electronic spectra, and their general conclusions may be stated very briefly. In the case of a molecule possessing a centre of symmetry, only the totally symmetrical vibrations will be excited in absorption from the ground state. For a molecule of lower symmetry, or for a molecule not in a vibrationless ground state we must apply the rule that "the transformation properties of the eigen-function of the vibration must be the same in the initial and in the final state". In other words the symmetry class must be the same in the initial and final states. These rules hold only when the molecule retains its symmetry group in going from the initial to the final state, and when it is assumed that the transition probability for the electronic jump does not depend on the nuclear motions of the molecule. When the latter condition is not fulfilled, the selection rules for the vibrational and electronic parts of the energy must be treated together. That is, the transformation properties of the complete electronic + vibrational initial state must be related in a definite way (defined by group theory) to those of the final state. Applications of these rules are now being made with considerable success to the simpler molecules, such as formaldehyde, and the methyl halides and they will obviously be of fundamental importance in all work of the future. It should be observed that the first two rules are an extension of the Franck-Condon principle for the intensities of vibrational transitions in diatomic molecules.

When we come to the rotational structure of the electronic bands, the difficulties are almost completely experimental. The resolving power required for all except the lightest molecules is so great as to be unobtainable unless in exceptionally well-equipped laboratories. Theoretically, it is little more complicated than in the vibrational spectrum when allowance is made for the fact that the moment of inertia may differ in the initial and final states. Unfortunately the phenomenon of predissociation is of common occurrence in polyatomic molecules and very frequently masks rotational structure. So far, only one or two polyatomic molecules have had the rotational structure of their electronic bands analysed. In particular we would mention the beautiful analysis of the formaldehyde bands by Dieke and Kistiakowsky[14]. Although several of the simpler molecules (NO_2, SO_2, ClO_2) exhibit discrete resolvable spectra, their analysis has as yet proved too difficult. However, now that the principles are well laid down, it should only be a matter of time before they succumb to the patience of the spectroscopist even as the very complex line spectra of atoms did at an earlier date.

Part 3. THE QUANTUM THEORY OF VALENCY

By W. G. Penney, M.A., Ph.D.

§ 1. INTRODUCTION

The first step towards an understanding of the mechanism of the chemical bond was made by Heitler and London in 1927 in their celebrated paper on the hydrogen molecule. Since then progress has been steadily maintained. It is the purpose of this article to give a short summary of the present position. Naturally, in the few pages at our disposal it is possible only to mention the results of a few of the more important papers, while the great majority cannot be discussed at all. The procedure we shall adopt is to take in succession various fields of the subject and give a brief description of the results which have been obtained in each. In § 2 we discuss calculations which have been made on the energies of formation of molecules; in § 3 we consider the shapes, sizes and vibration frequencies of molecules; in § 4 we deal with activation energies and reaction rates; and in § 5 we describe the calculations which have been made on the lattices of the alkali metals.

One of the most surprising features of the theory in its present form is the way in which results of extreme accuracy and results which can only be described as crude, are intermingled. To quote one example: the theory gives a very poor result (5 electron volts instead of 10 electron volts) for the energy of formation of the water molecule, yet it gives accurately the fundamental bending frequency of this molecule. Now vibration frequencies depend on the variations in the potential function of the molecule when small displacements are made from equilibrium. One is compelled to be sceptical of a calculation which gives the correct variations in the energy function for small displacements from the equilibrium position, but gives a very rough value for the function itself. To take another example: in the calculation of the activation energy of the ortho-para-hydrogen conversion, it is necessary to estimate the energy of the most stable configuration of the H_3 complex. If one takes only the "leading" terms of the energy, and this is what is always done in practice, a result in good agreement with experiment is obtained; but if one "improves" the calculations by putting in the "higher" terms (i.e. non-orthogonality and triple permutations), the corrections are so large that agreement with experiment is wrecked. Such a state of affairs is very discouraging. As the matter now stands, it is known from experience that for certain purposes, notably the calculation of activation energies, the numerical predictions of the theory can be relied on, but in other cases, such as energies of formation of molecules from atoms, the results will probably have only qualitative significance.

For an authoritative discussion of the present state of the quantum theory of valency the reader is referred to the article of Van Vleck and Sherman[1]. A much more elementary account is given in a monograph by the author[2].

§ 2. CALCULATIONS ON ENERGIES OF FORMATION OF MOLECULES

The most natural quantity to calculate from any theory of valency is the energy of formation of a molecule from atoms, and to a large extent the theory will be judged by its success in this direction. It is therefore disappointing to state that up to the present quantum mechanics has not proved capable of giving accurate values for the energy of formation of more than a few of the simplest diatomic molecules. The reason is that quantum mechanics deals with the interaction energies of electrons and nuclei, and therefore regards the energy of formation of a molecule as the *difference* between the interaction energy of the nuclei and electrons of the molecule and the sum of the interaction energies of the nuclei and electrons of the separate atoms. When the difference between two quantities, which must be evaluated separately, is small compared with either, the difference can easily be seriously in error. Perhaps an example will make the situation clear. The interaction energy of the electron and proton of the hydrogen atom is about 13·5 electron volts, and the energy of formation of the hydrogen molecule from atoms is about 4·5 electron volts. Hence the interaction energy of the protons and nuclei of the hydrogen molecule amounts to about 31·5 electron volts, and the sum of the interaction energies for the separate atoms is about 27·0 electron volts. The latter can be evaluated with great accuracy from theory, but the former offers more difficulty. An error of only 5 per cent in this quantity affects the value of the energy of formation of the molecule by 30 per cent. It is therefore a real triumph that the latest calculations[3] on the binding energy of the hydrogen molecule give a result whose error can be limited to a few hundredths of an electron volt. Unfortunately, the variational method by which this result was obtained is quite lengthy even for a molecule as simple as H_2, and its application to molecules with more than four or five electrons is too elaborate to be feasible.

Most calculations do not attempt to evaluate the energy of formation of a molecule from atoms as a difference of two large quantities, but make a direct attack at the difference itself. Molecular wave functions are constructed as the sum of products of atomic wave functions in such a way that the exclusion principle and the symmetry properties of the molecule are obeyed, and the energy of formation of the molecule is evaluated by the formula

$$E = \int H' \Psi^2 d\tau \qquad \ldots\ldots(1),$$

where H' is the perturbative part of the Hamiltonian. The expression for E is nothing other than the first term of an expansion which treats the molecular forces as a perturbation upon the system composed of non-interacting atoms. Hence, it is a good approximation only if the perturbation H' is not sufficiently strong to cause a considerable difference between the true wave function of the system and the "unperturbed" wave function Ψ. As a matter of fact, the molecular forces must often be large enough to cause appreciable distortion from Ψ, and when this is so conclusions based on (1) can have only qualitative significance.

Several authors have used (1) as the basis of computations on the energies of

formation of molecules. Thus Bartlett and Furry[4] have considered several states of the molecules Li_2 and Be_2; Ireland[5] has investigated a number of states of BeH; Coolidge[6] has carried through the calculations for the ground state of the water molecule; and Coulson[7] has worked out the energy curves for the equilateral H_3^+ ion. The results in some cases are satisfactory but in others are very disappointing.

With a view to testing how good are the approximations usually made in calculations on the binding energies of molecules, James[8] has made a careful investigation of several energy states of Li_2 and Li_2^+. He evaluates the energy from (1), with and without the customary approximations, and from a comparison of the results finds alarming discrepancies. Thus the modification caused by the inclusion of triple permutations and inner-shells is of the order of an electron-volt, and is therefore comparable with the energy of formation itself. Inner shells affect the result chiefly on account of their interaction with outer electrons of other atoms (with the understanding of course that each electron is taken together with its appropriate nuclear charge). Whether their influence is attractive or repulsive depends on the configuration of the electrons of the outer shells.

The conclusion which would appear inevitable in the light of the paper of James is that the calculation of the energy of formation of a molecule from atoms is an extremely difficult matter, and, except in the very simplest cases, is too complicated to be handled by present methods.

§ 3. THE SHAPE, SIZE AND VIBRATION FREQUENCIES OF A MOLECULE

A question of considerable interest is that of calculating the most stable configuration for the nuclei of a polyatomic molecule, or more briefly the shape and size of the molecule. Usually, of course, these are known already to the chemist, but this is not always the case. The problem can be summarized in a few words. It is required to evaluate the energy function of the molecule for all variations in the parameters which govern its configuration and then to determine the size and shape of the molecule by finding the minimum in the energy function. In spite of the formidable nature of this task, a certain amount of progress has been made. The usual line of attack is to obtain an expression for the energy function by as many simple approximations as possible and to see whether they all agree in predicting a particular structure. If they do, it seems likely that this result will persist in the accurate solution. By a process such as this, it has been shown, for example, that the four hydrogen atoms of CH_4 are arranged symmetrically around the carbon atom at the corners of a regular tetrahedron, that the molecule NH_3 has the shape of a rather flat pyramid with the nitrogen atom at the apex, and that the H—O—H angle of the water molecule is slightly in excess of 90°. While it cannot be said that demonstrations of this type are very satisfactory, at the moment it is, unfortunately, not possible to do anything better.

Let us suppose that by some method of approximation we have obtained an

expression for the energy of a molecule as a function of the coordinates determining its configuration. The shape of the molecule can now be found by discovering the position in which the energy has its minimum. A further calculation gives quite simply the *variations* in the energy for small displacements from equilibrium. From a knowledge of these the fundamental vibration frequencies can be calculated. Alternatively, the experimental evidence on the fundamental vibration frequencies permits these variations to be accurately determined. The procedure is therefore as follows. We obtain from some simple method of approximation the energy function of a molecule in terms of its internal coordinates and certain energy integrals. Approximate values for these latter can be obtained by direct numerical quadrature, and these are usually sufficiently good to enable the shape of the molecule to be determined. Having progressed so far, the next step is to obtain more accurate numerical values for the energy integrals; in other words, to adjust the values of the energy integrals which appear in the formula for the energy functions of the molecule (a formula, it will be remembered, which was obtained from an approximation) in such a way that the correct values are obtained for the heat of formation of the molecule and for the fundamental vibration frequencies. It is true that there are usually more energy integrals than there are equations to determine them from the experimental evidence, but in this case it is not a bad approximation to suppose that the ratios of the energy integrals are the same as those obtained by quadrature. By the process just described we have now constructed a formula for the energy function of the molecule which gives correctly the shape of the molecule, the heat of formation and the variations for small displacements from equilibrium.

As an illustration of the above we outline the theory of the ethylene molecule. For simplicity we shall adopt the viewpoint of the Slater-Pauling theory of directed valency, and shall make the usual assumption that the aggregate configuration of the carbon valency state is sp^3. The full calculations show that the most stable arrangement for the molecule is a plane one, with the H—C—C angles within a degree or two of 120°. We shall suppose that these angles are exactly 120°.

Denote the atoms of one CH_2 group by C_a, H_1 and H_2 respectively and similarly those of the other CH_2 group by C_b, H_3 and H_4. Then of the four "directed" wave functions of C_a, three are even for reflection in the plane of the molecule and have their maximum density in this plane, but the fourth is odd and has a node in this plane. The former can be represented.

$$\psi_i = (3)^{-\frac{1}{2}}\,[(2)^{\frac{1}{2}}\,\psi_a\,(\sigma_i 2p) + \psi_a\,(2s)], \quad (i = H_1,\ H_2 \text{ or } C_b),$$

while the latter can be represented by $\psi_a\,(\pi 2p)$. A similar set of orbital wave functions holds also for the electrons of the C_b nucleus.

The significance of the various symbols is as follows. The function ψ_i is a "directed" wave function, pointing towards the atom i. By $\psi_a\,(\sigma_i 2p)$ is meant the wave function of an electron on the nucleus C_a, the electron possessing zero component angular momentum with reference to the axis $C_a - i$ (the significance of σ_i), and being a $2p$-electron when the molecule is dissociated into atoms. It is usually assumed that a good approximation to the $\psi_a\,(\sigma_i 2p)$ wave function is that of a $\sigma 2p$

electron of the free atom. By $\psi_a(\pi 2p)$ is meant a $2p$-wave function on the nucleus C_a, the axis of symmetry of the wave function being perpendicular to the plane of the molecule. The meaning of $\psi_a(2s)$ is that it is a $2s$ wave function centred on the nucleus C_a.

Notice that the "hybrid" wave function ψ_i is very dense in the neighbourhood of the atom i and is small elsewhere. This comes about because $\psi_a(\sigma, 2p)$ and $\psi_a(2s)$ reinforce each other on the positive side of the $C_a - i$ axis, but partially cancel each other on the negative side. According to the directed valency theory, a bond between two atoms is represented by a pair of electrons, one on each nucleus, the two electrons having a resultant spin zero and the two electronic wave functions overlapping as much as possible. Thus the three wave functions ψ_i on each carbon nucleus are suitable for pairing off to form bonds. They give rise to the four C—H bonds and to the principal bond of the double bond. One unpaired electron remains on each carbon nucleus, and the remaining bond of the double bond is formed by the interaction of these two $\pi 2p$ electrons. These are the ones which compel the molecule to adopt a plane configuration.

If one of the CH_2 groups is rotated through an angle π about the C—C axis relative to the other CH_2 group, a system equivalent to the original one is obtained. Hence the energy function of the molecule has the property

$$W(\phi) = W(\phi + \pi),$$

where ϕ is the azimuth of one CH_2 group relative to the other. The simplest function which possesses this property is

$$W(\phi) = \text{const.} + A \cos 2\phi,$$

and closer investigation[9] shows that this is actually a good approximation. The quantity A can be expressed as an exchange integral which can be calculated.

Suppose now that one CH_2 group is twisted slightly out of its equilibrium position. The molecule will commence a "twisting" oscillation, whose fundamental frequency can be measured experimentally. From this information and the known dimensions of the molecule, the constant A can be determined. It is found that $A = 0{\cdot}55$ electron volt, in fair agreement with the value obtained by quadrature.

The energy needed to twist one CH_2 group through 90° with respect to the other is $2A$, and is therefore 1·1 electron volts. Now although this energy cannot be measured directly, it is possible to measure another quantity which should not be very different from it. If one of the H atoms of each CH_2 group is replaced by a $(COOCH_3)$ group, there are *cis* and *trans* forms for the resulting molecule, and they can be distinguished from each other as they have different chemical properties. Nelles and Kistiakowsky[10] have measured the energy of activation of one form to the other and find 1·15 electron volts. This is in excellent agreement with the value deduced from the twisting frequency in ethylene.

§ 4. ACTIVATION ENERGIES AND REACTION RATES

When two molecular systems collide they will in general separate again unchanged. Occasionally, however, the collision is sufficiently violent to cause a rearrangement of the atoms, and new products are formed. The violence of the collision necessary to cause a rearrangement to occur clearly depends on the relative orientation of the two systems on collision, and there will be one orientation which gives a path of least resistance. The energy of this collision is called the "activation energy" of the reaction. Unless the two colliding systems possess between them in available form energy equal to the activation energy, the collision is unable to bring about a rearrangement.* Most of the successful collisions in an actual experiment will follow the path of least resistance very closely, since the probability of a collision of a certain energy falls off very rapidly as the energy increases. For this reason, the energy of activation can be accurately measured.

A number of calculations have been made on the activation energies of chemical reactions. That successful results have been obtained is remarkable because activation energies as calculated by quantum mechanics are really second differences. Thus, the activation energy is the difference of the sum of the energies of formation of the separate molecules and the energy of formation of the complex; and the energy of formation of a molecule is the difference between the interaction energy of the nuclei and electrons of the molecule and the sum of the energies of interaction of the electrons and nuclei of the separate atoms. Thus, it is not surprising that the calculation of activation energies proceeds on somewhat empirical lines.

The method developed by Eyring and his co-workers[11] for the calculation of activation energies is based on electron-pairs, and is roughly as follows. Molecular energies are expressed in terms of two types of interactions, Coulomb and exchange. To take the simplest example, the potential function of the ground state of the H_2 molecule is written in the Heitler-London form

$$W = Q + J,$$

where Q is the Coulomb and J is the exchange interaction energy. From measurement on the band spectrum of H_2, values of W at various internuclear distances can be found. It is now assumed that the ratio (Q/J) is independent of the internuclear distance, and we shall later on state the method by which this constant ratio is determined. Since $(Q+J)$ and (Q/J) are now known for all internuclear distances, Q and J can be calculated separately as functions of the internuclear distance.

Let us now suppose that we have two molecular systems which can react when they are brought together. At first, when the systems are widely separated, the total energy, apart from thermal motions, can be expressed as ΣW_i, where W_i is the energy of formation of the system i. As the two systems are brought together they begin to interact and force is needed to prevent their flying apart. Under experimental conditions, of course, the thermal motions of the separate systems supply

* Neglecting the "tunnel effect", as is permissible for most purposes.

the energy needed to overcome this force. We require to calculate by how much the energy of the complex exceeds ΣW_i.

In order to proceed at all, the assumption is made that it is sufficient in the calculations to take account only of the bonds which are changed by the reaction. Since the pair theory interprets a bond as due to the interaction of a pair of electrons, the calculations can be restricted to $2n$ electrons, n being the number of bonds in the original system which are altered by the reaction. From the method of electron pairs, an expression is obtained for the energy of the complex in the form

$$W = C + \sum_{ij} Q_{ij} + f(J_{12}, \dots J_{ij}, \dots) \qquad \dots\dots(2),$$

where C is a constant representing the energy of the bonds which are not affected by the reaction, and Q_{ij}, J_{ij} are the Coulomb and exchange interactions of the electrons labelled i and j respectively. The form of the function f can be found in any particular example although, as is only to be expected, the complexity of the calculation from which it is obtained increases rapidly with n. The method of determination of Q_{ij} and J_{ij} is analogous to that used for the corresponding quantities in H_2. For example, in the reaction $H_2 + I_2 \rightleftharpoons 2HI$, there are H—H, I—I and H—I Coulomb and exchange interactions to consider. From measurements on the band spectra of the molecules H_2, I_2 and HI, the various quantities $(Q+J)$ can be found. The ratio (Q/J) is assumed constant and the same in all three cases, and this supposition enables each Q and J to be determined.

An expression has now been obtained for the energy of the complex in terms of the Q and J, which in turn can be expressed as functions of the internuclear distances governing the configuration of the complex. A convenient way of summarizing the situation is to take W as the dependent variable, and the internal coordinates of the complex as independent variables. Then W as given by (2) is the equation of a surface. We can if we wish measure W vertically upwards, and take the two horizontal coordinates as symbolic of all the independent coordinates of the complex. Each point on the surface represents a state of the complex; high regions correspond to unstable states, and lower regions to more stable states. We can describe the characteristic features of this surface by familiar words such as valleys, hills, passes and so on, and can represent the course of the reaction by the motion of a characteristic point on the surface.

At the beginning of the reaction, some coordinates will be very large and variations in them will affect but very little either the energy of the complex or the best values of the other coordinates. In other words, the reaction begins and ends with the characteristic point moving along the bottom of a valley. The course of the reaction is represented by the characteristic point following the bed of a valley up to a pass in the hills and then down the bed of another valley. The energy of activation of the reaction is the height of the pass above the bed of the first valley; similarly, the energy of activation of the reverse reaction is the height of the pass above the bed of the second valley. Clearly, the energies of activation of the forward and reverse reactions are not the same, but differ by the heat of the reaction.

We have still to describe how the ratio (Q/J) is determined. The method is to

consider (Q/J) as a parameter whose value is to be adjusted to give the correct activation energy to the ortho-para conversion in hydrogen. The same ratio is then taken over to all other interactions, except those involving a metallic bond.

Eyring and his collaborators have applied the above method to discuss the course of a number of reactions. Thus, among others, they have considered the ortho-para conversion, reactions between the halogens and hydrogen, between the halogens and the ethylene double bond, and the hydrogenation of the double bond in resonating systems. Agreement with experiment is extremely good and in no case does the discrepancy exceed a few kilo-calories. The results are in fact so satisfactory that some measure of confidence can be placed in predictions that have so far not been tested experimentally.

In view of the fact that the Eyring method gives such good results on systems which are altogether too complicated to attack by their "almost-rigorous" methods, Coolidge and James[12] have investigated the ortho-para conversion by the Heitler-London method, putting in all the terms. Their illuminating discussion does to some extent show how large terms ignored by the Eyring approximation practically balance out. On the other hand, there seems no reason why similar cancelling should occur in the case of other reactions. The conclusion of Coolidge and James is that the Eyring method must be considered as purely empirical. It is perhaps nearer the truth to regard it as an excellent interpolation formula.

If all the collisions which involve at least the activation energy caused the reaction to occur, a fairly good estimate of the rate of a reaction could be obtained from kinetic theory considerations. Although the rates calculated in this way are often about right, they are usually far too fast. To take account of this discrepancy the chemist introduces a steric, or orientation, factor, the purpose of which is to represent the fraction of the collisions involving at least the activation energy which are successful in causing a rearrangement. A theory which could calculate the steric factor of a collision would be of great value to the chemists. On physical grounds it is clear that if small changes in the relative orientation of the reactants at collision have the effect of raising considerably the energy needed for rearrangement, then the steric factor is very small; on the other hand, if the relative orientation has little effect, then most collisions involving at least the activation are successful, and the steric factor is nearly unity. Pelzer and Wigner[13] and more recently Eyring[14], have attempted calculations on the steric factors of simple reactions in the following way. First they construct the energy surface for the reaction and obtain the path of least resistance and the energy of activation. From a knowledge of the energy surface of the reaction it is possible to find the relation between variations from the activation energy of paths which deviate but little from the path of least resistance and the path itself. To find the steric factor it is necessary to average the number of collisions which involve at least the activation energy over all orientations, weighting each orientation with a probability factor that the collision has at least enough energy in excess of the activation energy to permit the reaction to occur at this orientation. It will be seen that calculations on rates of reaction involve *curvatures* of the energy surfaces. For this reason little significance can be attached to the

numerical results of the theory. Their value lies chiefly in allowing one to visualize the reaction and to see how one could calculate the rate accurately if one had sufficient patience to make the calculations rigorously in terms of electrons and nuclei.

§ 5. THE LATTICES OF THE ALKALI METALS

The theory of the lattices of the alkali metals has recently attracted considerable attention. Calculations have been performed with two objects. The first of these is to gain some understanding of the way in which the growth of the lattice proceeds. Closely connected with this problem is that of understanding why the heats of evaporation of very thin films of alkali metal are appreciably less than those of the bulk metal. The second is actually to calculate the Bloch wave functions of the alkali lattices with a view to using them in quantitative computations on various physical properties for which a general theory has already been developed.

Taylor, Eyring and Sherman[15] have made a very promising start on the former of these questions. They have applied the methods used by Eyring on activation energies to discuss the most stable arrangements of three, four, five, six, seven and eight sodium atoms, and have compared the energies of these systems with those of an equivalent number of diatomic sodium molecules. From their calculations it would appear that the formation of the unit cell proceeds as follows. Two diatomic molecules unite to form a square; a fifth atom adds along the cube edge; a sixth atom enters the lattice at the body centre of the cube; and the cell is completed by atoms adding at the remaining cube corners. The binding energy of even the eight-atom unit is only about the same as that of four diatomic molecules, while the bulk metal, of course, is appreciably more stable than the diatomic form. In order to obtain a unit approaching in stability that of the bulk metal it is therefore necessary to take one consisting of an appreciable number of unit cells. This is confirmed by the result that the nearest distance of approach of two atoms is 2·95 A. in the eight-atom cell compared with 3·7 A. in the bulk metal.

Experimentally, it is observed that finely divided particles of sodium are very much more active than the bulk metal. According to the present viewpoint, this is because the particles are so small that they do not contain enough atoms to give them a stability comparable with that of the bulk metal. A similar explanation will also account for the fact that the heats of evaporation of very thin films of alkali metal on quartz are only about one-tenth of those of the bulk metals.

Wigner and Seitz[16] have made an important contribution to the theory of the metallic state, and their results have been extended by Slater[17] and Millman[18]. The essence of their treatment is to replace the boundary conditions on the motion of an electron in the lattice by simpler conditions which actually permit a solution of the wave equation to be obtained. Now the alkali metals have a body-centred cubic structure. When lines are drawn connecting any atom with its eight nearest and six next-nearest neighbours, and planes are drawn bisecting these lines perpendicularly, the atom becomes surrounded by a truncated octahedron. The fundamental approximation on which the present section depends is that the potential

inside one of these polyhedra, or cells, is spherically symmetrical. This follows because each cell contains just enough negative charge to counterbalance its positive nuclear charge. If the cells were spherical, the potential at any point in one cell arising from the charge in another cell would be exactly zero. Since the shape of the cells is practically spherical and the distribution of the cells around any particular cell is highly symmetrical, the potential in any cell due to the charges in other cells is quite small and with good approximation has spherical symmetry. The potential inside a cell therefore arises mainly from the charge in the cell itself. On the average the charge distribution around a nucleus has spherical symmetry. Therefore it is a good approximation to assume that the potential field inside a cell is practically central.

The solution of the central field problem in quantum mechanics is well known. The wave equation separates immediately the solutions having the form $f(r)_l^m P(\theta,\phi)$. These solutions will not, however, satisfy the boundary conditions of the problem and it is therefore necessary to take a linear combination of them with arbitrary coefficients. The boundary conditions are that the solutions and their gradients join on smoothly at the boundaries of a cell to the solutions in the next cell. A non-vanishing solution of the wave equation which satisfies the boundary conditions is obtained in the usual way only by allowing certain values to the energy and momenta.

The results obtained from these calculations have been elegantly summarized by Slater and we refer the reader to his article for a detailed discussion of the method and the validity of the approximations introduced at the various points.

Part 4. THE THEORY OF PARAMAGNETISM AND RELATED PROPERTIES IN IONIC CRYSTALS

By R. Schlapp, Lecturer in Natural Philosophy, University of Edinburgh and W. G. Penney, 1851 Senior Exhibitioner, Trinity College, Cambridge.

§ 1. SCOPE OF THE ARTICLE

In this article we summarize the results of the theory of paramagnetism of free atoms and ions, as developed in the last few years chiefly by Van Vleck[1,2], and go on to outline the consequences of the hypothesis that each ion in a crystal is subject to an electrostatic field of definite symmetry due to its surroundings, in connexion with paramagnetic susceptibility, paramagnetic rotation, specific heats, and absorption spectra of crystals of salts of rare earths or iron group elements.

§ 2. MAGNETIC MOMENT OF A FREE ATOM

The carriers of the magnetism are the electrons of the atom or ion; their contribution is due partly to their intrinsic "spin", and partly to their orbital motion, the ratio of the magnetic moment to the angular momentum being twice as great

for the spin as for the orbital motion. The spin magnetic moment of a single atomic electron is $eh/4\pi mc$, the Bohr magneton, and will be denoted by β. The total angular momentum of the atom, which in the absence of external fields is conserved, is made up of orbital and spin contributions from each electron; but the magnetic moment of the atom will not in general be constant. In each energy state it will have a certain average value, equal to its projection on the constant total angular momentum, and will in addition have fluctuating terms. The average values and the fluctuating terms correspond respectively to the diagonal and non-diagonal terms of the matrix $\mathbf{M}$ which represents the magnetic moment.

The diagonal elements are readily calculated for a free atom with Russell-Saunders coupling, i.e. where the individual spin angular momenta and the individual orbital angular momenta of the atom build up to resultants $\mathbf{S}$ and $\mathbf{L}$ respectively. The angular momentum of the atom is $\mathbf{J}=\mathbf{L}+\mathbf{S}$, while its magnetic moment is $\beta(\mathbf{L}+2\mathbf{S})$. Hence the ratio of the projection of the magnetic moment on the angular momentum to the angular momentum is

$$\beta(\{\mathbf{L}+2\mathbf{S}\}.\{\mathbf{L}+\mathbf{S}\})/\mathbf{J}^2=\beta(3\mathbf{J}^2+\mathbf{S}^2-\mathbf{L}^2)/2\mathbf{J}^2.$$

Since $\mathbf{J}^2$, $\mathbf{L}^2$, $\mathbf{S}^2$, are diagonal matrices, with elements $J(J+1)$, $L(L+1)$, $S(S+1)$, this ratio may be written as $g\beta$, where

$$g=1+\frac{J(J+1)+S(S+1)-L(L+1)}{2J(J+1)}$$

is Landé's g-factor; in particular it has the value 1 for any singlet term, and the value 2 for any S term, but it may take values greater than 2 or less than 1. The non-diagonal elements can also be calculated quite simply, but we shall not give them here.

When the atom is subjected to a weak magnetic field, its orientation, hitherto arbitrary, becomes restricted to certain discrete directions such that the projection of $\mathbf{J}$ on the field direction, denoted by M, has any of the $2J+1$ values J, $J-1$, $J-2$, ..., $-J+1$, $-J$. Since the additional energy due to the magnetic field is given in first approximation by the scalar product of field and unperturbed average magnetic moment, the energy level (LSJ) will be decomposed into $2J+1$ nearly equidistant levels with displacements $Mg\beta H$. For the next approximation we must take into account the fact that the atom is distorted by the field, so that the fluctuating terms of the magnetic moment also make a contribution to the energy, which will however be proportional to the square of the field. This term can be regarded as representing the energy in the field H of the magnetic doublet, of strength proportional to H, induced by the field. The induced moment can be of either sign, and is small or large, according as the deformability of the atom, measured by the closeness of its energy levels, is small or large.

§ 3. ENERGY LEVELS, ZEEMAN EFFECT, AND MAGNETIC MOMENT

In the general case, where the angular momentum of the atom is no longer conserved owing to the disturbing influence of neighbouring atoms, we must fall back on the general theory to evaluate the magnetic moment. If we ignore the dia-

magnetic term (proportional to $\Sigma(x^2+y^2)$ for a field along z) the Hamiltonian for the atom is

$$\mathscr{H} = \mathscr{H}^0 - H\mathbf{M}.$$

Here $\mathscr{H}^0$ is the Hamiltonian for the atom in the absence of a magnetic field, and includes the perturbing influence of the neighbouring atoms as far as it can be represented by a constant electrostatic field, and $\mathbf{M}$ is the matrix representing the component along z (the field direction) of the magnetic moment. By the standard perturbation formula, the energy levels in the field are given by

$$W_{nn} = W_{nn}^0 - HM_{nn} + H^2\Sigma_l' \frac{|M_{nl}|^2}{W_{nn}^0 - W_{ll}^0} - \cdots \qquad \ldots\ldots(1).$$

Since the mean value of the magnetic moment along the field direction in the state (n, n) is given by $-\partial W_{nn}/\partial H$, we have for the moment

$$M_{nn} - 2H\Sigma_l' \frac{|M_{nl}|^2}{W_{nn}^0 - W_{ll}^0} \cdots.$$

The term independent of H corresponds to the permanent moment, and the term in H to the induced moment. The induced moment may be either in the field direction or in the opposite direction according to whether the effect of the quadratic term in (1) is to depress or raise the energy level.

A very important general result, of which we shall not give the proof here, is that an atom can have a permanent magnetic moment only if it is in a degenerate energy state. In other words, if W_{nn}^0 is a non-degenerate state, the mean moment M_{nn} vanishes, and there is no first order Zeeman effect. (The proof depends on the fact that the wave-function for a non-degenerate energy state has to be taken as real, and a real wave-function always gives zero angular momentum.)

§ 4. DEFINITION OF THE SUSCEPTIBILITY. CURIE'S LAW

The mean magnetic moment in the field direction, taking account of the different numbers of atoms in the various energy states, is

$$-\Sigma \frac{\partial W}{\partial H} e^{-W/kT} / \Sigma e^{-W/kT},$$

where W is the energy in presence of the magnetic field, and the sum is taken over all levels W (the subscript has been omitted). If this expression be expanded in powers of H, the term independent of H will vanish if the material has no permanent magnetization. The coefficient of H is what is usually called the susceptibility, and the higher terms give the behaviour in strong fields (saturation). If we rewrite (1) as $W = W_0 + HW_1 + H^2W_2 \ldots$ we find for the susceptibility χ, referred to a single ion, the formula

$$\chi = \Sigma e^{-W_0/kT} [W_1^2/kT - 2W_2] / \Sigma e^{-W_0/kT} \qquad \ldots\ldots(2),$$

which enables the susceptibility to be calculated as a function of temperature when the energy levels and their Zeeman effects are known.

To gain an insight into the meaning of this expression, consider a single magnetic level (assumed for simplicity to be doubly degenerate) having moment μ, lying at

a distance below all other levels which is large compared with kT. From the formula (2) the susceptibility will be $\mu^2/3kT$ (the factor $\frac{1}{3}$ occurring since μ is the whole moment, while W_1 is a component). If now the degeneracy of the level is removed, say by an electric field, it is fairly evident that, provided the separation is small, the susceptibility must remain the same. Suppose then that the level originally double separates into two with separation ΔW; these two levels will now have a matrix component μ of the magnetic moment between them, so that by (1) they will repel each other strongly by a quadratic Zeeman effect. The lower level gives an induced paramagnetic moment $2H\mu^2/3\Delta W$ and the upper an equal diamagnetic moment; but since the lower level is more densely populated than the upper in the ratio $e^{-\Delta W/kT} \fallingdotseq 1-\Delta W/kT$, the contribution of the pair of levels to the susceptibility is paramagnetic on the whole, and equal to $\mu^2/3kT$ as before. The result can be expressed by the statement that the "low-frequency" elements of the magnetic moment contribute in the same way to the susceptibility as the diagonal elements. The "high-frequency" elements, on the other hand, cause the low-lying levels to be depressed in a magnetic field, and therefore yield a paramagnetic contribution to the susceptibility, which may be shown to be independent of temperature.

When there are several levels with an over-all separation small compared with kT, lying far below all others, it can be shown that (2) leads to an expansion $\chi=\alpha+C/T+C\Delta/T^2+\ldots$, where $k\Delta$ is the energy difference between the "energetic" and "magnetic" mean centres of the low-lying group of levels[3]

$$k\Delta=\Sigma g_n W_{nn}{}^0/\Sigma g_n-\sum_{nl} g_n W_{nn}{}^0\,|\,M_{nl}\,|^2/\Sigma g_n\,|\,M_{nl}\,|^2.$$

Here g_n is the weight of the level $W_{nn}{}^0$. In first approximation this agrees with the Curie-Weiss form, $C/(T-\Delta)$ extended by the addition of a constant term α. Under certain conditions the constant Δ may be zero, in which case the susceptibility obeys the law of Curie; but in general the temperature dependence is expressible only as a power series in $1/T$.

In the rare earth salts it is reasonable to suppose as a first approximation that the spectrum of the ion in the crystal is very like that of the free ion; for the similarity in the chemical behaviour of the rare earths suggests strongly that the differences in the incomplete $4f$ shell do not make themselves felt outside the ion, and therefore that reciprocally the $4f$ electrons are only very little subject to external influences. It is true that the spectra of the free ions have not been analysed directly, but assuming Russell-Saunders coupling and Hund's rule for the multiplet of lowest energy arising from a given configuration (here f^x), we can predict the (LSJ) values of the lowest level, and hence the magnetic moment $g\beta\sqrt{\{J(J+1)\}}$ of the ground state. We may also assume that the multiplet intervals are wide compared with kT (except in the cases of Sm and Eu where a correction is needed), so that only the lowest multiplet level need be taken into account, and the susceptibility is $\chi=g^2\beta^2 J(J+1)/3kT$.

To facilitate comparison between the measured susceptibilities χ_m and this formula, it is usual to define an "effective magneton number" $\sqrt{(3kT\chi_m/\beta)}$ (which will vary slightly with temperature unless the substance obeys Curie's law accu-

rately), and compare it, usually at room temperature, with the theoretical value $g\sqrt{\{J(J+1)\}}$. This is of course rather a crude comparison; but it is found that the formula reproduces fairly well the values at ordinary temperatures of the susceptibilities throughout the rare earth group. It is not however surprising that this crude spectroscopic theory based on the assumption of free ions fails to represent the finer features of the temperature variation of the susceptibility, except in a few special cases. Usually the susceptibility appears to follow a law $C/(T-\Delta)$ over a fairly wide range, down to temperatures of 120° Abs. or less, below which "cryomagnetic anomalies" appear.

In the salts of the iron group elements the multiplets, which are in many cases known directly, are in general narrower, so the general formula (2) must be used to calculate the susceptibility. But in contrast with the case of the rare earth salts, the formula now gives results diverging considerably from the experimental values. This is to be expected, since the hypothesis of free ions is clearly a much poorer approximation in the elements of the iron group, with their comparatively exposed $3d$ electrons, than in the rare earths. It was noted by Bose, Sommerfeld, and Stoner, that fairly good agreement could be obtained with the observed magneton numbers by supposing that the orbital angular momentum of the atom does not contribute to the magnetic moment when it is in the crystal, so that the susceptibility has the value $4\beta^2 S(S+1)/3kT$.

It seems reasonable to suppose that both the cryomagnetic anomalies and the suppression of the orbital angular momentum are effects of the same nature, differing merely in degree, and are due to the electrostatic fields acting on the ion in the crystal. In the rare earth salts the effect of these fields will be to produce a slight Stark effect of the lowest level, along with a redistribution of the magnetic moment, which will affect the temperature dependence of the susceptibility especially at low temperatures. In the iron group the crystal field must be supposed strong enough to break down the coupling between the orbit and spin. Such a field will partially or completely remove the orbital degeneracy, and hence in general partially destroy the orbital contribution to the magnetic moment. One may picture the electric field as "quenching" the orbital angular momentum by causing it to execute large fluctuations so that it gives no contribution on the average to the magnetic moment. At the same time the freedom of the spin to orientate itself may be diminished in virtue of its coupling to the orbit. There is, however, no easy way of following out the effect of the crystal field in detail except by finding the energy levels of the ion in the field, and their Zeeman effects, and calculating the susceptibility from the general formula.

§ 5. CRYSTAL FIELDS AND THE INHOMOGENEOUS STARK EFFECT

The most important characteristic of the crystal field acting on an ion, apart from its absolute magnitude, is its symmetry type. The field on the individual ions does not of course necessarily have the symmetry of the crystal as a whole; thus for example the individual fields may have different orientations, thereby producing a higher total symmetry in the crystal.

The potential field, which is due chiefly to the immediate surroundings of the ion, may be supposed expanded in the region of the ion as a power series in (xyz), whose character will be largely determined by symmetry considerations. If the centre of the ion be taken as origin, and the electric force be supposed to vanish there, terms of the first degree will be absent; if the origin is a centre of symmetry, only even terms will be present. Thus if the surroundings of the ion have rhombic symmetry, the lowest terms that can occur are $Ax^2+By^2+Cz^2$; if the symmetry is that of the cube or octahedron, the lowest term is $D(x^4+y^4+z^4)$. It can be shown that the field of eight negative charges at the corners of a cube about the origin has the constant D negative, while six negative charges in octahedral arrangement at the face-centres give a positive D[4]. Fields of monoclinic or triclinic character could be represented by superposing a cubic and a rhombic part with different axes, and so on. The ideal calculation of the potential at the paramagnetic ion from a known crystal structure has not been carried out in any instance, so that the rapidity of convergence of the series representing the potential can only be estimated. In the explicit calculations made so far only the lowest terms giving the desired symmetry have been retained.

Once the appropriate potential function has been chosen, the energy levels and their Zeeman effects follow by the standard perturbation methods of quantum mechanics. The actual work is much facilitated by an investigation, due to Bethe[5], using the methods of group theory, on the decomposition of degenerate energy levels by fields of different symmetry. Typical results are that a field of cubic symmetry cannot decompose a P term (without spin); for a cubic field cannot distinguish between the three independent P wave functions which are proportional to x, y, and z; an axial field along z would decompose the P level into a single level, corresponding to the wave function z, and a doubly degenerate level, corresponding to the equivalent wave functions x and y. Again, a D term is decomposed by a cubic field into a doubly degenerate and a triply degenerate level, corresponding to the two types of independent harmonics of the second order, xy, yz, zx, and x^2-z^2, y^2-z^2. If spin is taken into account, half-integral values of the quantum number J will arise in atoms with an odd number of electrons; these have special properties from the point of view of group theory, which lead to the important result, due originally to Kramers[6], that an electric field can never wholly remove the degeneracy of the energy levels of an atom with an odd number of electrons, but must leave them with a double, or in general even, degree of degeneracy.

§ 6. MEAN SUSCEPTIBILITY OF PARAMAGNETIC CRYSTALS

It appears from X-ray evidence that in many highly hydrated crystals of paramagnetic salts the paramagnetic ion is surrounded by six oxygen atoms situated roughly at the corners of a regular octahedron, and giving rise to a crystal field which is predominantly cubic. This is almost certainly the case in the large class of simple hexahydrates, Tutton salts (double hexahydrates), and alums[7,8,9]. The arrangement of six oxygen atoms derived either from water molecules or, e.g., from

SO_4 groups around the paramagnetic ion seems to be a very general type of structure, and probably often occurs even when there are more or less than six water molecules to each paramagnetic ion, as for example in

$$Pr_2(SO_4)_3.8H_2O, \quad CuSO_4.5H_2O, \quad NiSO_4.7H_2O.$$

If this is so, one would expect that on the assumption of a cubic crystal field it should be possible to reproduce at any rate the main features of the temperature dependence of the susceptibility of many crystal powders. This has been done with satisfactory results for

$Pr_2(SO_4)_3.8H_2O$[10], $Nd_2(SO_4)_3.8H_2O$[10], $NiSO_4.7H_2O$[11],
$Co(NH_4)_2(SO_4)_2.6H_2O$[11], $KCr(SO_4)_2.12H_2O$[11], $CuSO_4.5H_2O$[12],

and several other compounds[13]. The dependence on temperature cannot be expressed by a simple formula; but in some cases, notably the Pr and Nd salts, it simulates a form $C/(T-\Delta)$ over a fairly wide range of temperature; this is, however, quite accidental. The order of magnitude of the crystal Stark effect which it is necessary to assume to account for the magnetic data varies from less than a volt in the rare earth salts to several volts in the salts of the iron group; in every case the sign of D which must be assumed corresponds with an octahedral arrangement of oxygen atoms round the metallic ion. In some cases in which the hypothesis of a cubic field is obviously inappropriate, a measure of success in accounting for the susceptibility has been obtained with other fields[14, 19].

At sufficiently low temperatures the susceptibility either becomes proportional to $1/T$, or tends to a constant finite value, according as the number of electrons in the ion is odd or even. This difference is strikingly illustrated by the behaviour at low temperatures of the susceptibility of salts so similar as $Pr_2(SO_4)_3.8H_2O$, $Nd_2(SO_4)_3.8H_2O$. The Nd ion has an odd number of electrons, so by the theorem of Kramers mentioned above the lowest level in the crystal field is degenerate; since it is magnetic, the susceptibility of the Nd salt is proportional to $1/T$ at low temperatures. On the other hand the lowest level of the Pr ion in the crystal turns out to be non-degenerate, and therefore non-magnetic, so that at low temperatures the susceptibility becomes independent of temperature, being due only to high-frequency elements of the magnetic moment.

§ 7. ANISOTROPIC SUSCEPTIBILITIES OF PARAMAGNETIC CRYSTALS

The hypothesis of crystalline fields of definite symmetry contains all that is necessary to account for the observed dependence of the susceptibility on direction in many non-cubic crystals. For if a crystal field of definite orientation and less than cubic symmetry acts on each ion in a rare earth crystal, producing a certain decomposition of its lower levels, the Zeeman effect of the spectrum, and consequently the magnetic moments of the levels, and the susceptibility, will in general be entirely different according to the direction of the magnetic field. In salts of the iron group the orbital moment may be supposed to be partially quenched; its remnants being anisotropic make the spin contribution to the susceptibility anisotropic as well.

Another feature for which it is possible to account is the observed variation with temperature of the directions of the principal axes of the susceptibility[15]. If for instance the crystal field had triclinic symmetry, consisting of a cubic part and a rhombic part superposed with different axes, the principal axes of the susceptibility would change with temperature.

As yet only a beginning has been made with explicit calculations. In the rare earth ions the problem is theoretically simpler than in the ions of the iron group, on account of the smaller Stark effects. But unfortunately only meagre crystallographic data, and up till quite recently, no magnetic data, have been available for single crystals of rare earth salts[16]. On the other hand reliable measurements of the principal susceptibilities of several iron group crystals have been available for some time. In the case of Ni salts, which show only about 2 per cent of anisotropy, it was found possible to obtain fair agreement between the calculated and observed principal susceptibilities on the assumption that the field acting on the ions is predominantly cubic, with a small superposed rhombic part having the same axes[11]. (The magnitude of a crystal field is supposed to be measured by the Stark effect it produces.) For the isomorphous cobalt salts, which have very much greater anisotropy, amounting to some 30 per cent, the calculation is more difficult, but it turns out[11] that sufficient anisotropy can be secured by assuming a crystal field of the same sort as in Ni, provided the whole Stark spectrum of the ground state be inverted in Co as compared with Ni. The lowest state of both ions is an F state, which is split by a cubic field into a single and two triply degenerate levels, the two latter being resolved completely by the superposed rhombic field. Clearly if the single level is lowest (which must be assumed to be the case in Ni) the anisotropy will be less than if the triple level is lowest. This inversion of the Stark spectrum of the orbital F state due to a cubic field as between Ni (d^8) and Co (d^7) was shown by Van Vleck[17] to hold generally for any pair of ions for which the number of $3d$ electrons adds up to 5, 10 or 15 (cf. the inversion of multiplets which occurs between pairs for which the number of $3d$ electrons adds up to 10; the present effect has no direct connection with this, but is clearly a phenomenon of the same type, where "missing" electrons play the part of electrons actually present, except that they occasion a change of sign). The success of this explanation, and the fact that the assignment to Ni of the pattern in which the single level is lowest corresponds to the field of six oxygen atoms round the Ni ion, must be regarded as evidence in favour of the type of field which has been assumed.

Similar calculations have been made on Cu salts[12], with satisfactory results. To explain the change with temperature of the directions of the principal axes of the susceptibility observed by Bartlett[15], a monoclinic field was assumed, consisting of a predominantly cubic field with a superposed rhombic field having only one of its axes coincident with a cubic axis. The amount of the variation in direction which could be obtained on this assumption was however not large enough.

§ 8. PARAMAGNETIC ROTATION

The theory of the rotation of the plane of polarization when light is propagated along the lines of force in a substance in a magnetic field (Faraday effect) has been worked out by Rosenfeld[18] for substances consisting of free paramagnetic atoms or ions. The rotation is found to consist of two parts, which show different dependence on frequency, and are respectively independent of, and in first approximation inversely proportional to, the absolute temperature; these are referred to as the "diamagnetic" and "paramagnetic" rotations. The diamagnetic part arises directly from the change produced by the magnetic field in the absorption frequencies and intensities for the two kinds of circularly polarized light. The paramagnetic part, with which we are here concerned, arises because the distribution of the atoms in the possible states depends on temperature through the Boltzmann factors. Thus the temperature dependence of the paramagnetic rotation, like that of the susceptibility, must be sensitive to crystalline Stark effect, which gives the possibility of correlating susceptibility and rotation measurements.

The question has been investigated theoretically by Kramers[19] and by Van Vleck and collaborators[20,21]. In the case of the rare earths it appears that if the energy changes due to the crystalline field are small compared with that of the transition responsible for the rotation, and with the multiplet separation of the ground state, the temperature dependence of the rotation and of the susceptibility will be the same. Explicit calculations[21] have been made for cerium, and the results compared with measurements[22] on tysonite (a mixed fluoride of cerium and other rare earth elements). Most of the rotation is due to an absorption band of Ce at about 40,000 cm^{-1}, and the crystalline Stark effect must be very much smaller than this. Each Ce ion is assumed to be subject to a rhombic field, but the orientations of these fields differ by rotations of 120° about the axis, so that the symmetry of the structure as a whole is hexagonal. This is in agreement with crystallographic evidence[23]. It is found possible to choose the parameters of the rhombic field in such a way as to account satisfactorily for the mean susceptibility of tysonite at all temperatures, and in addition for the measured values of the Verdet constants parallel and perpendicular to the crystallographic axis.

Kramers[19] has made similar calculations assuming each ion in tysonite to be subject to a trigonal field; this enables the mean susceptibility to be accounted for, but not the individual susceptibilities (or rotations)[21].

In the iron group the crystal field is large enough to produce a strong electric Paschen-Back effect of the ground state and also of excited states. The paramagnetic susceptibility and the paramagnetic rotation will therefore show quite different temperature dependence in the crystal; for the susceptibility is a property primarily of the low-lying levels, while the rotation involves frequency denominators corresponding to transitions to excited states. A possible exception, in which the rotation and the susceptibility should show the same temperature dependence, may occur in crystals of the iron group elements in which the ground state is an S state, and the ion is subject to a cubic field[20]. In this case the field is unable to decompose

either the S level, or the P level, transitions to which are involved in the rotation formula, except to a very small extent when orbit-spin coupling is taken into account. The experimental material is however too meagre to test this conclusion.

§ 9. SPECIFIC HEATS OF CRYSTALS AT LOW TEMPERATURES

Another phenomenon in which the crystalline field manifests itself is the temperature dependence of the specific heat of crystals of paramagnetic ions. The electronic contribution to the specific heat is

$$\frac{d}{dT}\left(\frac{\Sigma gWe^{-W/kT}}{\Sigma ge^{-W/kT}}\right),$$

where g is the statistical weight of the state of energy W; so that knowledge of the energy levels and their weights enables the contribution to the specific heat from this source to be calculated, and conversely, knowledge of the specific heat should give some information about the energy levels of the ion in the crystal field.

Recently Ahlberg and Freed[24] have investigated this problem in the salt $Sm_2(SO_4)_3.8H_2O$. In order to isolate the electronic specific heat in this substance, they compare its measured specific heat with that of the isomorphous $Gd_2(SO_4)_3.8H_2O$. In the latter the close conformity of the susceptibility to Curie's law down to liquid hydrogen temperatures shows that the spectrum of the Gd ion in the crystal consists of a single degenerate level lying far below all others; the Curie constant is consistent with a basic term $^8H_{\frac{7}{2}}$ which spectroscopic theory (Hund's rule) predicts. It is known that this term can be split only to a very small extent by the crystal field. A basic term $^8H_{\frac{7}{2}}$ is also in harmony with the observed inverse Zeeman effect at low temperatures.

It seems reasonable to suppose that any difference between the specific heats of the Gd and Sm ions in the crystal at all but very low temperatures (less than say 4° K.) is due entirely to the electronic contribution of the Sm ion. Ahlberg and Freed have determined the difference of the specific heats of the two crystals over the range 15–300° K. This difference rises from a very small value at 15° to some 2 cal. per half mole of either salt at about 110°, falls off again to 1·2 cal. at 220°, and rises rather sharply at higher temperatures.

From measurements of the absorption spectrum of crystals of $Sm_2(SO_4)_3.8H_2O$ Spedding and Bear[25] have deduced the existence of four low-lying levels situated at 0, 160, 188, and 225 $cm.^{-1}$ respectively. Since Sm has an odd number of electrons, these levels probably have statistical weight 2. Assuming these levels and weights, Ahlberg and Freed have computed the electronic specific heat of $Sm_2(SO_4)_3.8H_2O$ and compared it with that deduced from experiment. Good agreement was found up to about 150°, after which the calculated and observed values diverged considerably, probably owing to neglect in the calculation of all higher levels of the ion. Ahlberg and Freed conclude that it is not possible to fit the specific heat observations on the assumption that there are only the three low-lying levels which would be expected from a $^6H_{\frac{5}{2}}$ ground state.

At much lower temperatures than those considered by Ahlberg and Freed,

Kürti[26] has found considerable "anomalies" in the specific heat of $Gd_2(SO_4)_3 . 8H_2O$, using the term anomaly in the sense of a departure from Debye's T^3 law. These departures are due to the decomposition of the basic ${}^8H_{\frac{7}{2}}$ level of the ion in the crystal field; a splitting into eight equidistant levels with an over-all separation of the order of 1 cm^{-1} was found to give good agreement with the observed results. Since the Gd ion has an odd number of electrons, the levels in an electric field must all be at least doubly degenerate; this degeneracy must be supposed to be removed by magnetic interactions and exchange effects; for if for example the lowest level were doubly degenerate, the entropy at the absolute zero would be $k \log 2$, in contradiction with Nernst's Law. These anomalies in the specific heat are of course of great importance in connection with the method of Debye and Giauque for obtaining low temperatures by demagnetization.

§ 10. ABSORPTION SPECTRA OF PARAMAGNETIC CRYSTALS

Perhaps the most direct evidence of the existence of a crystalline Stark effect is afforded by the absorption spectra of crystals of the rare earth and iron group elements, and ultimately it must be possible to correlate these spectra with the magnetic and other data. Extensive measurements of absorption spectra and their Zeeman effects[25,27,28] have already been made; in the case of the rare earths the spectra at ordinary temperatures consist of fairly sharp lines, and at low temperatures some of the lines are as sharp as those of the emission spectra of gases. In crystals of the iron group elements (e.g. chrome alum) the lines at ordinary temperatures are diffuse, but most of them become sharper as the temperature is reduced. The relative intensities of the lines also undergo remarkable changes as the temperature falls; some of the lines become fainter and ultimately disappear, and new lines may appear. Most of these changes can be understood in a general way in terms of the Boltzmann distribution of the atoms in the various states, and it should be possible on this basis to deduce the positions of the low-lying levels of the ions in the crystal. Up to the present, however, the interpretation is not very certain, and the problem is a difficult one, especially in the rare earths, where the spectra contain an enormous number of lines. In cases where an energy level diagram has been deduced for the ion in the crystal, puzzling divergences from the behaviour to be expected on spectroscopic theory sometimes appear. For instance, in Sm Spedding and Bear[25] find a group of definitely more than three low-lying levels within a range of 300 cm^{-1} while the most that could arise from a basic term ${}^8H_{\frac{7}{2}}$ in the free ion would be three. A further difficulty is of course that the spectrum of the free ion is known for none of the rare earths. In the case of the iron group elements, the situation is more favourable in this respect, but on the other hand the crystal fields are stronger, and there must be strong perturbations between different terms. Some beautiful measurements have been made by Spedding and Nutting[28] on the multiplet at 6700 A. in the absorption spectrum of chrome alum, including its Zeeman effect. Again it is not immediately obvious how these measurements fit into the scheme which was successful in accounting for the susceptibility of chrome

alum[11]. The direct spectroscopic test is of course far more exacting than that depending on susceptibility measurements; so far no absorption spectrum has been interpreted quantitatively on the theory of the crystal Stark effect.

§ 11. CONCLUSION

While the earlier experimental results were often inadequate and unsystematic, much material has been accumulating in recent years in a form suitable for testing the theory of the crystalline field. What is required is complete measurements of all the effects discussed above for crystals whose structure is known, containing a paramagnetic ion whose spectrum is known in the gaseous state. Ideally the problem of correlating the effects quantitatively by the idea of the crystal field is soluble, taking into account if necessary the exchange forces between the ions, which give rise to a quasi-ferromagnetism [20]; the actual calculations may however become very involved. As yet only a beginning has been made, and a great deal of work, both experimental and theoretical, remains to be done in this field.

REFERENCES

(This list is not intended to be complete, it contains only those papers in which details of the work mentioned in the report can be found.)

Part I

(1) Andrade. *Report on Progress in Physics*, p. 269 (1934).
(2) Bethe. *Phys. Rev.* **47**, 633 (1935).
Oliphant, Kempton and Lord Rutherford. *Proc. roy. Soc.* A, **150**, 241 (1935).
(3) Wigner. *Phys. Rev.* **43**, 252 (1933).
(4) Heisenberg. *Z. Phys.* **77**, 1 (1932); **78**, 156 (1932); **80**, 587 (1933); *Rapports du VII*ième *Conseil Solvay* (Paris, 1934).
(5) Majorana. *Z. Phys.* **82**, 137 (1932).
(6) Chadwick and Goldhaber. *Nature*, Lond., **134**, 237 (1934). *Proc. roy. Soc.* A, **151**, 479 (1935).
(7) Chadwick. *Proc. roy. Soc.* A, **142**, 1 (1933); *Rapports du VII*ième *Conseil Solvay* (Paris, 1934).
Wigner. *Z. Phys.* **83**, 253 (1933).
Wick. *Z. Phys.* **84**, 799 (1933).
Massey and Mohr. *Proc. roy. Soc.* A, **148**, 206 (1935).
Bethe and Peierls. *Proc. roy. Soc.* A, **149**, 176 (1935).
(7a) Fermi. *Phys. Rev.* **48**, 790 (1935).
(8) Bethe and Peierls. *Proc. roy. Soc.* A, **148**, 146 (1935).
(9) Feenberg. *Phys. Rev.* **47**, 850 (1935).
(10) Young. *Phys. Rev.* **47**, 972 (1935).
(11) Gamow. *The Constitution of Atomic Nuclei* (1931). Oxford.
(12) Goldhaber. *Proc. Camb. phil. Soc.* **30**, 561 (1934).
(13) Amaldi, d'Agostino, Fermi, Rasetti and Segrè. *Proc. roy. Soc.* A, **149**, 522 (1935).
(14) Bethe. *Phys. Rev.* **47**, 747 (1935).
Perrin and Elsasser. *J. Phys. Radium*, **196**, 194 (1935).
(15) Pauli. *Rapports du VII*ième *Conseil Solvay* (Paris, 1934).

(16) Bethe and Peierls. *Nature*, Lond., **133**, 532 (1934).
(17) Ivanenko. *Nature*, Lond., **129**, 798 (1932).
(18) Fermi. *Z. Phys.* **88**, 161 (1934).
Wick. *Accad. R.C. Lincei*, **19**, 319 (1934).
(19) Konopinski and Uhlenbeck. *Phys. Rev.* **48**, 1 (1935); **148**, 107 (1935).
(20) Tamm and Ivanenko. *Nature*, Lond., **133**, 981 (1934).
(21) Heisenberg. *Zeeman Verhandelingen*, p. 8 (1935). The Hague.
(22) Wick. *R.C. Accad. Lincei*, **21**, 170 (1935).

Part 2

(1) Collins and Price. *Rev. sci. Instrum.* **5**, 423 (1934).
(2) Mecke. *Hand- und Jahrbuch der Chemischen Physik.* **9** (1934).
(3) Adel and Dennison. *Phys. Rev.* **44**, 99 (1933).
Teller. *Hand- und Jahrbuch der Chemischen Physik.* **9** (1934).
Johnson and Dennison. *Phys. Rev.* **48**, 868 (1935).
(4) Sutherland and Dennison. *Proc. roy. Soc.* A, **148**, 250 (1935).
(5) Dennison. *Rev. mod. Phys.* **3**, 280 (1931).
(6) Tisza. *Z. Phys.* **82**, 48 (1933).
(7) Placzek. *Leipzig. Vortr.* (1931); *Handbuch der Radiologie*, **6** (1934).
(8) Lennard-Jones. *Trans. Faraday Soc.* **30**, 70 (1934).
(9) Bethe. *Ann. Phys.*, Lpz., **3**, 133 (1929).
(10) Mulliken. *Phys. Rev.* **43**, 279 (1933).
(11) —— *Phys. Rev.* **41**, 751 (1932), and also Ref. (8).
(12) Eastwood and Snow. *Proc. roy. Soc.* A, **149**, 434 (1935).
(13) Herzberg and Teller. *Z. Phys. Chem.* B, **21**, 410 (1933).
(14) Dieke and Kistiakowsky. *Phys. Rev.* **45**, 4 (1934).

Part 3

(1) Van Vleck and Sherman. *Rev. mod. Phys.* **7**, 167 (1935).
(2) Penney. *The Quantum Theory of Valency* (1935). Methuen.
(3) James and Coolidge. *J. chem. Phys.* **1**, 825 (1933).
(4) Bartlett and Furry. *Phys. Rev.* **37**, 507 (1931).
Furry and Bartlett. *Phys. Rev.* **38**, 1615 (1931); **39**, 210 (1932).
(5) Ireland. *Phys. Rev.* **43**, 329 (1933).
(6) Coolidge. *Phys. Rev.* **42**, 189 (1932).
(7) Coulson. *Proc. Camb. phil. Soc.* **31**, 244 (1935).
(8) James. *J. chem. Phys.* **2**, 794 (1934); **3**, **8** (1935).
(9) Penney. *Proc. roy. Soc.* A, **144**, 166 (1934); *Proc. phys. Soc.* **46**, 333 (1934).
(10) Nelles and Kistiakowsky. *J. Amer. chem. Soc.* **54**, 2208 (1932).
(11) Sherman, Sun and Eyring. *J. chem. Phys.* **3**, 49 (1935). This paper gives the references to all the earlier papers of Eyring and his co-workers.
(12) Coolidge and James. *J. chem. Phys.* **2**, 795 (1934).
(13) Pelzer and Wigner. *Z. phys. Chem.* B, **15**, 445 (1932).
(14) Eyring. *J. chem. Phys.* **3**, 107 (1935).
(15) Taylor, Eyring and Sherman. *J. chem. Phys.* **1**, 68 (1933).
(16) Wigner and Seitz. *Phys. Rev.* **43**, 808 (1933); **46**, 509 (1934).
Wigner. *Phys. Rev.* **46**, 1002 (1934).
Seitz. *Phys. Rev.* **47**, 400 (1935).
Wigner and Bardeen. *Phys. Rev.* **48**, 84 (1935).
(17) Slater. *Rev. mod. Phys.* **6**, 209 (1934).
(18) Millman. *Phys. Rev.* **47**, 286 (1935).

PART 4

(1) VLECK, J. H. VAN. *The Theory of Electric and Magnetic Susceptibilities* (1932). Oxford.
(2) STONER, E. C. *Magnetism and Matter* (1934). Methuen.
(3) GORTER, C. J. *Arch. Mus. Teyler*, **7**, 183 (1932).
(4) —— *Phys. Rev.* **42**, 437 (1932).
(5) BETHE, H. *Ann. Phys.*, Lpz., **3**, 133 (1929).
(6) KRAMERS, H. A. *Proc. Acad. Sci. Amst.* **33**, 962 (1930).
(7) HOFFMANN, W. *Z. Krystallogr.* **78**, 279 (1931).
(8) BEEVERS, C. A. and LIPSON, H. *Z. Krystallogr.* **83**, 123 (1932).
(9) VEGARD, L. and ESP, E. *Ann. Phys.*, Lpz., **85**, 1152 (1928).
(10) PENNEY, W. G. and SCHLAPP, R. *Phys. Rev.* **41**, 194 (1932).
(11) SCHLAPP, R. and PENNEY, W. G. *Phys. Rev.* **42**, 666 (1932).
(12) JORDAHL, O. M. *Phys. Rev.* **45**, 87 (1934).
(13) JANES, R. B. *Phys. Rev.* **48**, 78 (1935).
(14) PENNEY, W. G. *Phys. Rev.* **43**, 485 (1933).
(15) BARTLETT, B. W. *Phys. Rev.* **41**, 818 (1932); **44**, 687 (1933).
(16) FEREDAY, R. A. and WIERSMA, E. C. *Physica*, 's Grav., **2**, 575 (1935).
(17) VLECK, J. H. VAN. *Phys. Rev.* **41**, 208 (1932).
(18) ROSENFELD, L. *Z. Phys.* **57**, 835 (1930).
(19) KRAMERS, H. A. *Proc. Acad. Sci. Amst.* **35**, 1272 (1932); **36**, 17 (1933).
(20) VLECK, J. H. VAN and PENNEY, W. G. *Phil. Mag.* **17**, 961 (1934).
(21) VLECK, J. H. VAN and HEBB, M. H. *Phys. Rev.* **46**, 17 (1934).
(22) BEQUEREL, J., HAAS, W. J. DE and HANDEL, J. VAN DEN. *Physica*, 's Grav., **1**, 383 (1934).
(23) OFTEDAL, I. *Z. phys. Chem.* B, **5**, 272 (1929); **13**, 190 (1931).
(24) AHLBERG, J. E. and FREED, S. *J. Amer. chem. Soc.* **57**, 431 (1935).
(25) SPEDDING, F. H. and BEAR, R. S. *Phys. Rev.* **42**, 58 (1932); **42**, 76 (1932); **44**, 287 (1933); **46**, 308 (1934); **46**, 975 (1934).
(26) KÜRTI, N. *Z. phys. Chem.* B, **20**, 305 (1933).
(27) SPEDDING, F. H. *J. chem. Phys.* **1**, 144 (1933).
(28) SPEDDING, F. H. and NUTTING, G. C. *J. chem. Phys.* **2**, 421 (1934); **3**, 369 (1935).
(29) KRAMERS, H. A. *Physica*, 's Grav., **1**, 182 (1934).

ATOMIC PHYSICS

By N. FEATHER, Ph.D.

§ 1. INTRODUCTION

This, the second report in the present series on the physics of the atom[1], is intended to deal chiefly with those contributions to experimental knowledge in this field which were made during 1934. At the time of writing, however, seven months more have passed—quite a considerable interval when a rapidly developing subject is in question. New knowledge has been gained, and in some cases the outlook of December 1934 is already out-of-date. Whilst keeping, as far as possible, references to this recent work brief and subsidiary to the main theme, such changes in viewpoint have not been allowed to pass unnoticed. If this involves the anticipation of next year's report, in some degree, it is obviously dictated by the march of events.

Two reports on atomic—more particularly nuclear—physics, of outstanding authority, have been published during the year. They deal, respectively, with the proceedings of the Solvay Conference at Brussels in October 1933 and with those of the International Conference which met in London and Cambridge very nearly a year later. These reports are considerably more detailed than anything which can be attempted here; they have been freely used, however, and may be further referred to, for information concerning the earlier work.

§ 2. THE SYSTEM OF THE ELEMENTS

From the point of view of physics, as is well known, the classification of elements by chemical properties, which issues essentially in the Periodic Table of Mendeléeff, is a classification of atomic species according to nuclear charge number ("atomic number", Z). In course of time it has been ascertained that the species which occur in the earth's crust possess all integral values of nuclear charge between 1 and 92 (in terms of the proton charge), with the possible exception* of charges 61, 85 and 87 units[2]. Moreover, the recently discovered neutron represents the species for which $Z=0$. As far as can be judged, most of these elements are stable, on the other hand some (the heavy radioelements other than uranium and thorium) are, more or less rapidly, being lost by spontaneous transformation and so depend for their present existence upon earth on their continued production in the radioactive decay of parent species of greater stability (ultimately, uranium or thorium). As

* If the β-activity of neodymium[3] be regarded as established, then also is the formation, as subsequent product, of a nucleus having $Z=61$. Similarly, if the work of Gueben[4] be accepted, the element 87 is the product of a very rare (frequency 10^{-4} to 10^{-10}) α mode of disintegration of mesothorium 2.

regards abundance, enormous differences—represented by factors of 10^6 or more—occur as between one stable species and another. That in three cases this abundance should be below the level of detection need, therefore, occasion no surprise.

The recognition of the phenomenon of isotopy* among elements opened a new field, which, through the work chiefly of Aston, is already almost exhausted, in so far as concerns the determination of the mass numbers of existing stable species. At the end of 1934 only eight mass numbers (5, 8, 57, 105, 191, 193, 194, 195) had not with certainty been allotted to any element, whereas four elements (palladium, iridium, platinum and gold) remained for which attempts at analysis had so far failed.† The discovery of artificial radioactivity[8] strengthened rather than lessened the conviction that the great majority of stable atomic species had already been discovered. Figure 1 exhibits the reason for this conclusion very forcibly as far as the light elements are concerned. Here known nuclei‡ are represented on a two-dimensional diagram in which axes of Z and $A-Z$ are employed. (On current views the number of protons in the nucleus is thus plotted against the number of neutrons.) Stable nuclei are indicated by full circles, unstable nuclei by open circles.§ It will be observed that the former lie in a well-defined region, flanked on either side by the open circles. Since only integral values of the "variables" are physically significant, it is obvious that the possibility of discovery of further stable species in this range is very restricted indeed. When a similar diagram is constructed to include the heavier elements, also, an interesting feature is revealed. For odd values of Z never more than two full circles appear, and, when two are present, they correspond to adjacent even values of $A-Z$, being separated by a single blank space in the array of points—cf. the full circles at (18, 17) and (20, 17) in figure 1. What amounts precisely to this regularity was noted long ago by Aston. Now the discovery of artificial radioactivity has strengthened the conviction that this feature, also, reflects a fundamental nuclear instability in the "missing" isotopes of odd atomic number, for in some cases, at least (Cu^{64}, Ga^{70}, Br^{80}, Ag^{108} and In^{114}), it appears reasonably certain that the hitherto unknown isotope of intermediate mass is one of the spontaneously radioactive products of the artificial disintegration of known stable nuclei. Similarly, the view is gaining ground[9] that the radioactivity of potassium is to be ascribed to K^{40}, present in small amount, rather than to either of the known isotopes K^{39} or K^{41}. Likewise, with rubidium, Rb^{86} may well be responsible.

At the present time, amongst the naturally and "artificially" radioactive elements—other than the products of the three main disintegration series—samarium is the only element known to emit heavy particles in the course of its

* Whereby, for a given charge number (or, for identical chemical properties), more than one mass number (A) is possible for an existing species.

† Three of these elements (palladium, platinum and gold) have recently been analysed[5] with the result that mass numbers 105, 194, 195 have now been allotted. Mass number 57 belongs to iron[6] and, almost certainly, 193 to iridium; similarly, there is growing evidence both for Li^8 and for Be^8[7].

‡ The figure is complete to July 31, 1935.

§ The neutron is here shown as a stable nucleus. This, however, may not be the case. On the other hand, Be^8 is shown as unstable and Be^{10} as stable: the reverse may possibly be correct.

transformation. No definite information is available concerning the isotope here responsible, but certain displacement anomalies in hyperfine structure patterns have led to the suggestion[10] that Sm^{152} (or Sm^{151}, present in small amount) is the species involved. It is also suggested that an $\alpha\beta\beta$ sequence might correlate the β activity of neodymium and the extreme scarceness of "illinium" (61) on earth with the above-mentioned samarium activity; thus,

$$Sm \xrightarrow{\alpha} Nd \xrightarrow{\beta} Il \xrightarrow{\beta} Sm.$$

The discovery of "artificial" radioactivity places the "natural" radioactivity of potassium, rubidium, neodymium and samarium somewhat in a new light. These radioactive species now appear as survivors of an earlier chaos only on account of their extremely long life time. Quantum theory shows that, for a given energy of disintegration (α or β), such long lived radioelements may be expected when the change of spin between initial and final nuclei is no longer small. It is quite likely that no other explanation will be needed here.

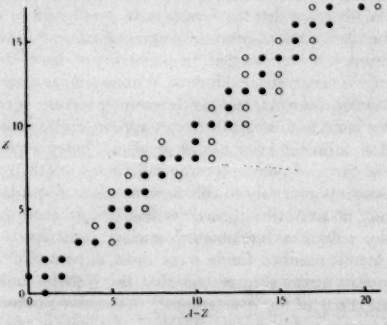


Figure 1.

As concerns the radioelements of the main series, it has often been pointed out that, for formal completeness, a fourth such series would be required. Only α particles and β particles being emitted during the disintegrations, the mass numbers of all the members of a series may be represented by a single expression, $4n+m$, with n variable. Four sets of series are then possible, with $m=0, 1, 2, 3$, respectively. For the thorium series $m=0$, for the uranium series $m=2$ and for the actinium series it is becoming increasingly certain[11] that $m=3$. Now, members of the hypothetical series for which $m=1$ have been produced (see below) as "artificial" radioelements by bombarding thorium with neutrons. The distinction between natural and artificial radioelements here appears at its flimsiest extreme.

If the mass numbers of almost all stable species have at last been determined by mass spectrum analysis there remains the equally important and more difficult

problem of fixing the exact masses. The mass of the nucleus (or of the neutral atom) has to be measured in terms of some convenient unit, e.g. one-sixteenth of the mass of the neutral atom of oxygen. When this is done by the usual methods, a series of intermediate comparisons is frequently involved in which errors may be cumulative. On the other hand certain mass differences may be very accurately deduced from data on artificial disintegrations[12]. As the latter became more exact, definite discrepancies in accepted values began to be detected. Artificial disintegration data relative to the light nuclei have now been reviewed by Oliphant, Kempton and Rutherford[13] and by Bethe[14] and independent self-consistent sets of masses assigned. New determinations by Aston, with an improved positive ray collimator[15], show that these doubtless possess a great measure of truth.* The previously accepted values, those suggested by Oliphant and Bethe, and the new values of Aston are compared in table 1.

Table 1

Nuclear species	Masses of neutral atoms				To obtain nuclear mass, subtract
	Mass spectrograph 1934	Oliphant	Bethe	Aston, 1935	
n^1	—	1·0083	1·0085	—	—
H^1	1·00778	1·0081	1·00807	1·0081	0·00055
H^2	2·01363	2·0142	2·0142	2·0148	0·00055
H^3	—	3·0161	3·0161	—	0·00055
He^3	—	3·0172	3·0170	—	0·0011
He^4	4·00216	4·0034	4·0034	4·0041	0·0011
Li^6	6·0145	6·0163	6·0161	—	0·0016
Li^7	7·0146	7·0170	7·0169	—	0·0016
Be^9	9·0155	9·0138	9·0135	—	0·0022
B^{10}	10·0135	10·0143	10·0146	—	0·0027
B^{11}	11·0110	11·0110	11·0111	—	0·0027
C^{12}	12·0036	12·0027	12·0037	12·0048	0·0033
C^{13}	—	—	13·0069	—	0·0033
N^{14}	14·008	—	14·0076	—	0·0038
N^{15}	—	—	15·0053	—	0·0038
O^{16}	16·0000	16·0000	16·0000	16·0000	0·0044
O^{17}	—	—	17·0040	—	0·0044

Most of the original discrepancies were due to an error in the He^4/O^{16} ratio, of which one effect, it may be remarked, was the assigning of too low a mass to H^1. Temporarily, at any rate, this was not without its compensation, for the initial impetus in the successful search for H^2 was provided chiefly by the conflict between this low value and the chemical atomic weight of elementary hydrogen[16]. It is now clear that this conflict was, to a very large extent, illusory.

During 1934 new methods have been devised for the determination of nuclear spin and magnetic moment. These methods will be referred to in detail later. The

* When the rest of Aston's apparatus has also been improved this work is to be continued.

earlier work on this subject, and the necessity for this data for a complete description of nuclear fields of force and of the interaction between nucleus and extranuclear electrons in an atom, have already received attention in last year's *Report*[17]. It will suffice here to remark that the magnitudes of these two vector quantities, I and μ, when expressed in fundamental units, are the one (I) integral* and the other (μ) non-integral—a distinction which also holds as between the magnitudes of the more familiar scalar quantities, nuclear charge and mass, already discussed. It is agreed, then, that these four quantities, together, are necessary for the complete description of nuclear fields; whether or not they are sufficient has been a matter for frequent speculation[18]. Very recently, however, a quantitative explanation has been offered[19] of certain interval anomalies in hyperfine structure patterns[20] in terms of a nuclear electric quadripole moment. Whilst it is unlikely that this is a common attribute of nuclei, it appears probable that in certain cases (Eu^{151} or Eu^{153} and Lu^{175}), when a spin axis occurs, the nett positive charge on the nucleus is distributed about this axis as a unique axis of revolution. In such cases the distribution of charge is no longer spherically symmetrical, even for the nucleus in its ground state. Whilst it is generally obvious, therefore, that there is plenty of scope for further experimental work directed towards the determination of nuclear constants—and any theory of nuclear constitution must evidently depend upon them—this report must now cease from generalizations and be concerned in greater detail with the trend of that work, in a number of particular directions, during 1934.

§ 3. THE SEPARATION OF ISOTOPES

The importance of determining the nature of disintegration phenomena, as well as certain of the constants described in § 2, for samples of material in which the atomic nuclei are all of one kind has resulted in a renewal of interest in the separation of isotopes—a subject which at one time appeared to have but a dwindling academic importance. The various methods of separation—or enrichment—of general applicability were set out as early as 1923 by Aston[21]. Most of them required that the material used should be in the form of a gas or, at least, be susceptible of distillation. In all cases, for the enrichment to be at all considerable, an enormous number of repetitions of the chosen process was imperative. It was obvious that the positive ray analysis of the mass spectrograph afforded the only method of complete separation, but at that time its success appeared scarcely possible. It is just this method which is now being satisfactorily applied to a number of elements. That it has proved successful is as much due to the fact that, frequently, extremely small amounts of the separated material suffice for the experiments which are contemplated, as to the advance recently made in the design of high intensity instruments of analysis. Yet the new experimental arrangements are, in themselves, of sufficient interest to warrant description here. First, however, reference should be made to the diffusion method of Hertz, which has recently been applied to the elements hydrogen[22] and neon[23]. Because of the very high ratio of the masses,

* The chosen unit, in this case, being $\frac{1}{2}\frac{h}{2\pi}$, where h is Planck's constant.

hydrogen constitutes a special case—for which even chemical and electrochemical methods provide sufficiently powerful discrimination[24]—and no difficulty was discovered in obtaining spectroscopically pure H^2 by repeated diffusion. But with neon, also, Harmsen, Hertz and Schütze were able to separate the heaviest isotope, Ne^{22}, almost completely from its lighter congeners. A fresh modification of the original method is now being tried, with considerable success[25].

The positive ray methods have been developed by Smythe, Rumbaugh and West[26] and Oliphant, Shire and Crowther[27], respectively. These two groups of workers utilized quite different means of attaining the intense ion currents through the analysing fields which are necessary for the collection of satisfactory samples of material. Smythe, Rumbaugh and West arranged to collect about 10 per cent of the positive ions emitted from an extended source (30 cm^2 area), after the ions of a single mass had been brought together in a short focal line and admitted through a narrow slit into the receiving cup. With potassium they were able to focus a collecting current of 100 microamperes of positive ions and with lithium one of several microamperes, so that, even in the latter case, one microgram samples of the separated isotopes were frequently obtained. Oliphant, Shire and Crowther prepared lithium samples, weighing one or two tenths of a microgram, using correspondingly smaller effective currents. These currents, however, represent a very small fraction of the total emission (several hundred microamperes) provided by their positive ion source, a platinum filament coated with a mixture of oxides and carbonates, and in their apparatus, also, the area upon which the final deposit was collected was not markedly smaller than that of the source. To a certain extent, therefore, efficiency of operation was sacrificed to ease of construction—but, with lithium, at least, the yields obtained were quite sufficient for the disintegration experiments which were important at the time. Figure 2 *a* and *b* represents diagrammatically the essential features of the apparatus of Smythe, Rumbaugh and West. It should be noted that the two portions of the figure are drawn to quite different scales and that they constitute sections of the assembled apparatus by planes at right angles. Figure 2 *a* shows the arrangement of source and the electrodes for the accelerating electric field, which also serves to collimate the beam. The main difference of potential is applied between the cylindrical emitter *A* (only partly covered with active substance, as shown) and the bevelled slit *B*. Refinements in collimation are obtained by giving suitable potentials to the wire *C* and the bars *D*, *D'*. A parallel beam of ions then issues from *D*, *D'* of cross-section roughly 25 cm. by 6 mm. It is this beam of ions (or the constituents of given e/mv) which is brought to a focus in an area about 2 by 6 mm. after traversing a magnetic field of appropriate intensity generated between parallel pole pieces of the shape shown in figure 2 *b*. To a first approximation (in practice corrected for edge effect) the field boundaries for entrance and emergence of the beam are the circular arcs *OB*, *OA*, centres *D* and *C*, respectively. When the axis of incidence of the beam is parallel to *OC*, as indicated, the focal line is perpendicular to the figure and passes through *C* (optimum adjustment of field). Ions of slightly different e/m values are then focussed in parallel lines intersecting *OC*.

Oliphant, Shire and Crowther used two arrangements in each of which crossed electric and magnetic fields were employed in opposition. This combination of fields is essentially a velocity selector for charged particles; it is effective for mass analysis if equally charged particles of different masses enter it along the axis after having fallen through the same difference of potential. In one apparatus this potential difference was about 800 V. and the cross-section of the beam in the fields about 2 cm. by 2 mm., in the other apparatus an accelerating voltage of roughly 10,000 V. was so distributed as to give a focussing effect (the focal spot being just in front of the collecting surface), and the channel through the fields was consequently of larger area of section.

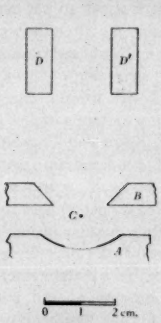


Figure 2 (*a*).

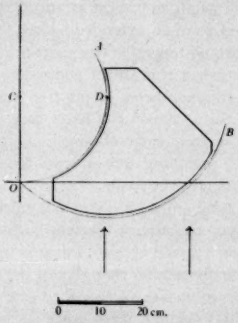


Figure 2 (*b*).

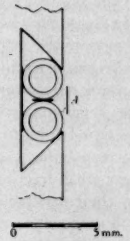


Figure 3.

A method of isotope separation of considerable interest, but very restricted applicability has been developed by Szilard and Chalmers[28]. Fermi had found that when neutrons are employed to produce artificial radioactivity in heavy elements, very frequently, the radioactive atom which is formed is isotopic with the atom originally affected (see below). The problem was to concentrate the active isotope so produced. With elements of the halogen group this problem proved particularly easy of solution. A liquid organic compound was used and a trace of the free halogen was added. When, during irradiation of such a mixture, a radioactive nucleus is formed by capture of a neutron, the impact must, in general, be sufficient to remove the halogen atom or ion from the molecule in which it is bound. A complete molecule will tend to re-form, but in most cases an atom of the originally free halogen will be recaptured rather than one of the much smaller number of such atoms freed in the process of activation. After irradiation, therefore, the greater bulk of the activity may be separated by any chemical process which removes the free halogen present. The experiments of Szilard and Chalmers confirmed this conclusion, and the method itself has been extended to other elements, involving the use of inorganic compounds, with and without protective additions of the element to be activated—either in elementary or ionic form[29].

§ 4. NEW METHODS OF DETERMINING NUCLEAR SPIN I* AND MAGNETIC MOMENT μ

Before 1934 almost all determinations of nuclear spin were made either from band spectrum observations—relative to elementary diatomic molecules—or from the analysis of the hyperfine structure of atomic spectral lines. The latter method provided data from which, in addition, the nuclear magnetic moment might be calculated in favourable cases. These two methods have already been described in this series of *Reports*[30]. In one particular case, also, direct determinations of nuclear magnetic moments had already been effected during 1933. Stern and his collaborators[31] deduced the magnetic moments of proton and deuteron by determining directly the resultant molecular magnetic moments for the ortho and para modifications of light and heavy hydrogen, respectively. In this way they were able to analyse the effective moment of the ortho-hydrogen molecule into its rotational and nuclear components, utilizing the fact that in para-hydrogen (nuclear spins anti-parallel) the molecular magnetic moment is entirely rotational. The data for ortho- and para-deuterium were treated in an analogous manner. The original Stern-Gerlach deflection method was employed, although the field gradients necessary were considerably greater than those previously required for the determination of atomic moments. The resultant molecular moment of ortho-hydrogen at liquid air temperatures, for example, is about 1/300 of the magnetic moment of the atom of silver. During 1934, however, another modification of the Stern-Gerlach method was successfully applied to the determination of nuclear magnetic moments. The suggestion had been made by Breit and Rabi[32] in 1931. These authors do not determine nuclear moments directly; their method is rather to be related to the hyperfine structure method of determination, in that both I and μ may be obtained from the final analysis if certain auxiliary data are available. For the hyperfine structure method the magnetic splitting of atomic energy levels is fundamental: the number of components may give information concerning I, their separations ($h\Delta\nu$) on the energy scale are determined directly by μ. Likewise the atomic beam method of Breit and Rabi provides data from which the magnetic splitting, $\Delta\nu$, may also be obtained. Thus, if a complete quantum specification of the ground state of the atom is available, and if I be known, μ may be calculated, as before. The chief advantage lies in the fact that in the method of atomic beams the splitting of the energy level is directly investigated, whilst in the former method, in general, it may be deduced only after an analysis of the hyperfine structure patterns of suitable spectral lines. On the other hand, when this analysis has been carried out, even for one line, the magnetic splitting of two energy levels has been determined. It has been a very necessary and useful check upon theory to examine for consistency the values of μ calculated from the magnetic splitting of different energy levels of the same atom.

* More exactly, throughout this section, I is used to denote the quantum number specifying nuclear angular momentum.

The method of Breit and Rabi utilizes the fact that in weak magnetic fields, when the coupling between nucleus and electronic system has not completely broken down, space quantization of an atom results in $(2I+1)(2J+1)$ permitted directions with respect to the field.* Corresponding to these directions the component along the field of the magnetic moment of the atom (in all practicable cases due almost entirely to the electron shells) takes on values which differ one from another and depend upon the field strength H, being given precisely in terms of the non-dimensional parameter $eH/2\pi mc.\Delta\nu$. As the field strength increases these values tend to coalesce in groups of $(2I+1)$ and only $(2J+1)$ essentially distinct orientations remain possible. For certain intermediate fields, also, when $I \geqslant 1$, a subsidiary degeneracy appears, the values in question becoming equal in pairs. In particular, pairs of equal zero values of the longitudinal component of atomic magnetic moment may occur. A powerful modification of the original method has been based upon this peculiarity[33]: it is most useful with mixed elements in which different isotopes are characterized by different values of I and μ[34]. Whatever modification is used, however, the aim is to measure, by means of deflection of a beam of atoms in an inhomogeneous magnetic field, the various components of magnetic moment along that field. From the number of maxima in the deflection pattern (or in the graph of undeflected intensity against field strength) I is determined; from the separations of these maxima (in units of length, or of field strength) $\Delta\nu$ may be calculated, and so μ, as above. As regards the technical standpoint, the problem has been to produce sufficiently weak fields possessing sufficiently high gradients. One solution of this problem[35] is illustrated in figure 3. It embodies the great convenience that, owing to absence of iron, fields and gradients may be calculated from the magnetizing current and the geometry of the apparatus, alone. Currents of as much as 250 A. (maintained for 8 hours) were passed through two parallel copper tubes of 2·4 mm. outside diameter, spaced by 0·05 mm. mica. Water was circulated through the tubes, which in one experiment[36] were more than 60 cm. long. The cross-section of the beam is represented in the figure by the line A. Over this range the field and its gradient were both constant within a few units per cent. Other arrangements of field have since been employed[37]. In one particular, however, the results so far obtained by the method of atomic beams have been incomplete. Hyperfine structure analysis has shown that, for certain nuclei, the axis of magnetic moment is opposite in direction to that of the mechanical moment of the nucleus. For these nuclei μ is said to be negative. The experiments of Stern and his colleagues on the hydrogen isotopes, together with those so far carried out by the method of Breit and Rabi, are entirely inconclusive in the matter of the sign of μ. By an elaboration of the method, however, Rabi[38] hopes to remove this ambiguity in certain cases.

The progressive magnetic uncoupling of the nucleus from the extra-nuclear system in fields of increasing strength has already been discussed. One result of the complete uncoupling in strong fields is that in the Zeeman effect in such fields

* J is the quantum number specifying the resultant angular momentum of the extra-nuclear system.

each main component of the pattern consists of $(2I+1)$ equally spaced subcomponents, due to the independent orientation of the nucleus in the field. If the separation is not too small this splitting of the Zeeman components may be utilized for a direct determination of I[39]. Usually, however, it is impossible to observe the splitting, directly; then, sometimes, observations on a closely related phenomenon may be made which supply the data required. It may be shown that when the resonance spectrum of an atom is excited by plane polarized radiation then the percentage polarization of the resonance radiation which is emitted must depend upon the value of I. The amount of this polarization may be increased by the application of a magnetic field in a direction parallel to that of the electric vector in the incident light. Numerically, the increase depends upon $H/\Delta\nu$, where H is the field strength and $\Delta\nu$ is the hyperfine structure separation of the upper level involved in the electron transition. The rate of increase of polarization with field strength, then, depends ultimately upon μ. In 1932 Heydenburg, Larrick and Ellett[40] determined the percentage polarization of the sodium-D resonance radiation, excited by plane polarized light in the absence of an applied field, and attempted to deduce the nuclear spin of sodium (Na^{23}) from their results. However, the value which they obtained did not agree with that determined by other methods. Later the matter was investigated in great detail theoretically by Breit[41]. It appears from his discussion, first, that no essential contradiction was involved in the original data, and, secondly, that because of experimental difficulties the method may best be applied to cases where I is known—when μ may be evaluated by observing the variation of percentage polarization with applied field, as above described. In that way the magnetic moments of the nuclei of sodium[42] and caesium[43] have recently been determined.

The ratio of the magnetic moments of proton and deuteron have been calculated by Kalckar and Teller[44] from the observation of Farkas, Farkas and Harteck[45] that the rate of reconversion of para-hydrogen into normal hydrogen by paramagnetic oxygen is sixteen times greater than that of the similarly catalysed change from ortho-deuterium to the equilibrium molecular mixture. The value so obtained[46] (4 : 1) is in good accord with other determinations.

These new methods, then, have added greatly to the possibilities of determining the important rotational characteristics of nuclei, angular momentum (spin) and magnetic moment. Already for more than half of the known stable nuclei of odd mass number* the nuclear spin has been determined by one method or another. Frequently, several methods have given completely concordant results. For more than thirty of these nuclei the magnetic moment is known with more or less certainty. This is a considerable achievement, but it seems inevitable that much more information will be obtained in the future. Closer examination of hyperfine structure patterns, from the point of view of the validity or failure of the interval

* Stable species of even mass number, with the exception of H^2, Li^6, B^{10} and N^{14}, have even atomic number also. For such atoms (A and Z even) it has generally been assumed that the nuclear spin is zero. Although the general body of evidence points to this conclusion, it must still be emphasized that band spectrum observations—which provide by far the most critical test for $I=0$—are, for these species, almost entirely lacking.

rule, may—as previously mentioned—provide evidence for an electric quadrupole moment in certain nuclei, and an investigation, experimental and theoretical[47], of the isotope displacement frequently observed in such patterns may provide independent data concerning, not only the isotopic constitution[48] of certain elements, but also the relative binding energies (mass defects) of the nuclei concerned[10].

§ 5. NUCLEAR TRANSFORMATION BY FAST PARTICLES

The nuclear transformations discussed in last year's *Report*[49] were almost entirely of a single type. They were transformations in which two colliding and two (or three) resultant nuclear particles are involved. They may be referred to as processes of capture disintegration. Typical examples are represented by the nuclear schemes:

$$ {}_7N^{14} + {}_2He^4 \rightarrow {}_8O^{17} + {}_1H^1 \qquad \ldots\ldots(1), $$

$$ {}_3Li^7 + {}_1H^2 \rightarrow 2\,{}_2He^4 + {}_0n^1 \qquad \ldots\ldots(2). $$

Before taking up the discussion of transformations of this type from the stage in which it was then left, it will be convenient first to discuss certain other types of transformation for which the evidence became more definite during 1934.

First, there are processes of non-capture excitation. That is to say that the struck nucleus, or the colliding particle, is raised to a state of higher energy at the expense of the energy of motion of the particle, but no synthesis takes place. Subsequently a quantum of radiation is emitted from the excited system. It is the detection of this nuclear "resonance" radiation, in circumstances which preclude any other interpretation, which constitutes the sole evidence for transformations of this type. It appears probable that they occur in certain cases when α particles and neutrons are used as projectiles. When α particles are used it is well known that, with many substances, capture disintegration takes place as a result of which the final nucleus is left in an excited state[50]. The problem is to discriminate between the subsequent γ radiation from such excitation and that produced as the result of any other (e.g. non-capture) process. If, however, γ radiation is found when the energy of the incident α particles is already so small that capture disintegration (with excitation) is impossible, excitation without capture must be postulated to account for the facts. In this way[51] it has been concluded that the nuclei Li^7[52], N^{14}[53], F^{19}[54] and Al^{27}[55] may be excited by α particle bombardment. With other nuclei, too, a similar effect may be produced, but the data are not sufficiently extensive or accurate for any conclusion to be drawn except in the most obvious cases. Similarly, when neutrons are employed as projectiles, there is as great, if not greater, difficulty in interpretation. For, monoenergetic[56] neutrons have not generally been used, and almost always there has been no information regarding the energy threshold for the capture transformations which certainly occur. Despite these facts, however, it appears probable that the γ radiation observed by Lea[57] to be produced by the passage of neutrons through many substances, is to be ascribed to the inelastic scattering of the neutrons in close collision with atomic

nuclei. From other experiments entirely Dunning[58] has also concluded that inelastic scattering is of frequent occurrence when fast neutrons and heavy nuclei are involved.

The second type of process now to be considered is that of simple capture. This is merely a process of synthesis: the incident particle and the struck nucleus coalesce and the excess mass and kinetic energy is liberated as a quantum (or quanta) of radiation. For this process to occur the emission of radiation must obviously take place in a time comparable with that during which the interaction between the particles remains large when the collision is elastic. It has generally been concluded, on this basis, that the probability of simple capture processes must always be small[59]. On the other hand, the experiments of Fermi and his colleagues have shown that with neutrons as projectiles simple capture occurs with a great number of elements. Moreover, when the neutron velocity is first reduced (as by multiple scattering in hydrogenous substances) the cross-section for capture may become very great indeed[60]. These experiments will be described more fully in § 6. Evidence for the simple capture of protons by nuclei was first advanced by Cockcroft, Gilbert and Walton[61], who have studied in detail the reaction

$$ {}_6C^{12} + {}_1H^1 \rightarrow {}_7N^{13} \qquad \text{.....(3).} $$

Again, however, the final nucleus is "artificially" radioactive. Other cases of simple capture of protons (or of other particles) will doubtless be discovered as more exact analysis becomes possible of the transformation products set free in different experiments. Already, from the observations of Crane, Delsasso, Fowler and Lauritsen[62] it appears probable that the reactions

$$ {}_4Be^9 + {}_1H^1 \rightarrow {}_5B^{10} \qquad \text{.....(4),} $$

$$ {}_5B^{11} + {}_1H^1 \rightarrow {}_6C^{12} \qquad \text{.....(5),} $$

take place under suitable conditions, since γ rays of approximately the predicted energy have been observed. Similarly, the high energy γ rays shown by the same authors[63] to be produced when lithium and fluorine are bombarded by protons are now more probably ascribed to the reactions

$$ {}_3Li^7 + {}_1H^1 \rightarrow {}_4Be^8 \qquad \text{.....(6),} $$

$$ {}_9F^{19} + {}_1H^1 \rightarrow {}_{10}Ne^{20} \qquad \text{.....(7),} $$

than to those previously discussed[64]. This conclusion follows directly from the observations of Hafstad and Tuve[65]. These authors showed that the artificial radioactivity produced in carbon by reaction (3) and the γ-ray emission excited in lithium and fluorine, as above described, were similar in a most important respect. In each case a resonance phenomenon occurred, since only protons of definite energies were at all effective in producing the transformation. The excitation functions for the reactions involved in the previous interpretations of the emission were known to be without appreciable indications of resonance.

From the theoretical point of view the process of radiative capture has been re-discussed by Breit and Yost[66] for protons and, amongst others, by Bethe[67]

for neutrons. It appears that, in general outline, the observed phenomena may be qualitatively explained. In the theoretical treatment, also, the question of possible resonance occupies a prominent position.

Non-capture disintegrations may be considered third in this survey. By non-capture disintegration is meant essentially an inelastic encounter between a fast particle and an atomic nucleus in which some of the original energy of motion is used in splitting the nucleus into two or more fragments, which appear as products of the reaction. This type of process is perhaps less well attested than any at present under consideration. It is probable[58] that the deuteron may be disintegrated by bombardment by fast α particles*

$$ {}_1H^2 + {}_2He^4 \rightarrow {}_1H^1 + {}_0n^1 + {}_2He^4 \qquad \text{.....(8),} $$

but, as concerns neutron-produced disintegrations[69], the position is still rather indefinite. Unless, however, production of the nucleus He^5 be assumed, the triple disintegration fork observed by Chadwick, Feather and Davies[70] must obviously be interpreted in terms of the non-capture process

$$ {}_6C^{12} + {}_0n^1 \rightarrow 3\,{}_2He^4 + {}_0n^1 \qquad \text{.....(9).} $$

At one time the balance of evidence appeared to favour this view, but now that calculations have been carried out with the new atomic masses of Oliphant and Bethe the other alternative appears rather to be gaining ground. On a non-capture basis the energy of the responsible neutron is advanced from 4·5 to about $9{\cdot}5 \times 10^6$ electron volts by the change—already approaching the energy limit for the neutrons from polonium-beryllium—whilst, assuming capture, the atomic mass of the hypothetical He^5 becomes 5·0096, in close agreement with the value 5·010 predicted by Oliphant, Kempton and Rutherford[13] from general considerations.

The discussion of capture disintegrations has been reserved for the fourth, and last, subdivision of this section. Transformations of this type constitute by far the greatest fraction of those known to occur when atomic nuclei are bombarded by α particles, protons, deuterons or neutrons. One example of α-particle transformation,

$$ {}_{13}Al^{27} + {}_2He^4 \rightarrow {}_{14}Si^{30} + {}_1H^1 \qquad \text{.....(10),} $$

has been considered in detail in § 8 of last year's *Report*. Similarly, § 6 of that *Report* discussed the effects obtained with protons and deuterons as projectiles in deuterium, lithium and boron, respectively. During 1934 a great deal of additional data was collected concerning these and similar cases of transformation. Moreover, in that year it first appeared[71] that unstable nuclei might result from the disintegrations in question. Previously it had been supposed that the final nucleus was formed either in the ground state—when it was stable indefinitely—or, in a certain fraction of cases, initially with excess energy, which was subsequently liberated as a quantum (or quanta) of γ radiation. The latter process might be characterized by a half-value period, if required, but this was obviously so short ($\sim 10^{-10}$ sec.) that it did not enter directly into any experimental effect. During

* A theoretical calculation relevant to this and other reactions in which the deuteron may be disintegrated without capture has recently been given by Oppenheimer[68].

1934 it appeared that, in a great many cases, unstable nuclei were formed for which the half-value period for electron (or positron) emission might be of the order of hours or days. As far as possible, however, these cases, and the properties of the unstable species produced, will be discussed in the following section on "artificial radioactivity".

The new data[72] concerning disintegrations produced by α particles are important because they provide information not only for the possible energies of excitation of the product nuclei (from one standpoint, the relative separations of the negative energy levels for α particles in those nuclei), but also because they determine the conditions of entry of the incident particle into the struck nucleus. For all energies greater than a given energy, E_0, and for values of the energy neighbouring upon discrete energies, E_1, E_2, etc., lower than E_0, the α particle appears to have reasonable probability of entering into the nuclear structure. E_0 is referred to as the energy corresponding to the top of the potential barrier for the given nucleus as against α particles, whilst E_1, E_2, etc., fix the positions of the α-particle resonance levels (nuclear levels of positive energy) in the barrier[73]. Obviously, information of both types is relevant to any theory of nuclear structure which may be advanced; for that reason, in particular, other possible sources of this information are very desirable. As concerns the nature of the potential barrier, only, confirmatory and additional evidence may perhaps be expected from a more detailed study of neutron-produced disintegrations. This possibility will be referred to later. Similarly, when a particular product nucleus may be obtained as the result of more than one process of transformation, further opportunities may be afforded of examining the excited states of that nucleus[74].

Transformations produced by protons and deuterons—more especially the latter—have been investigated in great detail by many workers. The disintegration particles have been analysed in Cambridge by Oliphant and Rutherford[75], Cockcroft[76], Dee[77] and their associates; in Berkeley, California, by Lawrence[78] and his co-workers. The γ rays emitted in the same disintegrations have been studied by Crane, Delsasso, Fowler and Lauritsen (Pasadena)[79]. The Cambridge workers, operating with low- or medium-voltage tubes, have concentrated attention on the lightest elements; Lawrence, on the other hand, having deuterons of more than 2×10^6 eV. energy available, has been able to investigate the corresponding reactions with substances of higher atomic weight. Perhaps the most interesting case, because of the initial difficulties of interpretation, is that of beryllium. It seems likely that all the reactions (11) to (16)

$$ {}_4Be^9 + {}_1H^1 \rightarrow {}_4Be^8 + {}_1H^2 \quad \text{.....(11),} $$

$$ {}_4Be^9 + {}_1H^1 \rightarrow {}_3Li^6 + {}_2He^4 \quad \text{.....(12),} $$

$$ {}_4Be^9 + {}_1H^2 \rightarrow {}_4Be^{10} + {}_1H^1 \quad \text{.....(13),} $$

$$ {}_4Be^9 + {}_1H^2 \rightarrow {}_4Be^8 + {}_1H^3 \quad \text{.....(14),} $$

$$ {}_4Be^9 + {}_1H^2 \rightarrow {}_3Li^7 + {}_2He^4 \quad \text{.....(15),} $$

$$ {}_4Be^9 + {}_1H^2 \rightarrow {}_5B^{10} + {}_0n^1 \quad \text{.....(16),} $$

actually occur, as well as reaction (4), previously discussed. With the heavier

elements (e.g. nitrogen, sodium and aluminium) it appears that neutrons, protons and α particles may be obtained when deuterons are employed for the bombardment. As already described, a portion of the γ radiation observed when protons are used must certainly arise in simple capture processes; on the other hand more frequently with deuterons the γ-ray energies may be satisfactorily identified with the excitation energies known, from other observations, to characterize the residual nuclei in reactions such as (16).

As regards neutron-provoked capture disintegrations, in this section we shall be concerned chiefly with those cases known to occur because of the direct observation of the products of transformation. For experimental reasons their number is relatively small; until recently the expansion chamber provided the only means of investigating them[73]. Now, in certain cases, two* other methods have also proved successful. Chadwick and Goldhaber[81] have employed an ionization chamber and proportional amplifier to study the effect of slow neutrons on a number of light elements. With lithium and boron a copious† production of charged particles was observed, together with a much smaller, though definite, effect with nitrogen. For the success of the method complications due to the projection of nuclei by elastic collision must obviously be avoided: it is useful, therefore, only when slow neutrons are employed. In that case it may not be used when endothermic processes are in question, nevertheless a sufficient number of exothermic processes occur with the lighter elements to make its application worth while. Because of the long range (5·5 cm. in air) of the disintegration particle from lithium, it was possible to study this case in some detail. The effective transformation was shown to be

$$ {}_3Li^6 + {}_0n^1 \rightarrow {}_2He^4 + {}_1H^3 \qquad \text{.....(17).} $$

These effects, with lithium and boron at least, have since been examined by Fermi and his collaborators in a similar manner[82]. Taylor and Goldhaber[83] have applied the method of the photographic emulsion[84] to their further elucidation. They introduce salts containing the element to be investigated into the gelatine emulsion of a photographic plate and expose this to a source of neutrons for a convenient time. After development the plate is examined through the microscope for "tracks" of silver grains indicating the previous passage of charged particles. The search for types of emulsion most suitable for this work has been described by Taylor[85]. So far only the disintegrations of lithium and boron have been studied, the most frequent process with boron being shown to be

$$ {}_5B^{10} + {}_0n^1 \rightarrow {}_3Li^7 + {}_2He^4 \qquad \text{.....(18).} $$

Obviously the method is equally satisfactory when fast neutrons or slow neutrons are employed.

* The scintillation method of investigation has also been applied to the problem by Matzner[80].

† The yield in these reactions is so large that the most sensitive and convenient methods of neutron detection now make use of them. A small ionization chamber, coated inside with boron (or lithium) and surrounded on all sides by paraffin (about 6 cm. thick), is connected in the usual way to a proportional amplifier system. The neutrons, slowed down first of all by collisions in the paraffin, liberate heavy particles by disintegration in the chamber. The bursts of ionization so produced are recorded in any convenient manner.

Work with the expansion chamber has not progressed very rapidly, for lack of fresh substances which may be introduced into the chamber in the gaseous state.* Carbon, nitrogen, oxygen, fluorine, neon and argon have been examined[87], but technical difficulties have prevented the extension of this list to other elements, e.g. silicon, phosphorus, sulphur and chlorine. Nevertheless, when more intense sources of neutrons are available, this method of investigation is likely to assume a fresh importance. For then it will be possible to make a statistical analysis of the energies with which the α particle is emitted relative to the recoiling nucleus in such reactions as

$$ {}_8O^{16} + {}_0n^1 \rightarrow {}_6C^{13} + {}_2He^4 \qquad \text{.....(19)}, $$

and, when due allowance has been made for the distribution of energy amongst the neutrons responsible, this information will refer directly to the penetrability of the potential barrier of C^{13} as against α particles[72]. Moreover, in the example chosen, C^{13} is a relatively rare atomic species, concerning which such information might be difficult to obtain by other means.

In the matter of neutron sources many times more intense than those which may be prepared from polonium (or radon) and beryllium, Kurie has advanced considerably, by employing the neutrons obtained from reaction (16) when the energy of the incident deuterons is of the order of 2×10^6 eV. He has described[88] the results of expansion chamber experiments with nitrogen under these conditions. Obviously, with sources only a few times more intense still, it will be possible to collect disintegration data at a very profitable rate. Kurie[89] has attempted to interpret the data so far obtained from a novel point of view. He seeks to explain the large values of the kinetic energy loss almost invariably found with fast neutrons† by the assumption that disintegration occurs in two stages. First, the neutron is captured, thus radiating a large fraction of its energy, then, after a finite time, the new nucleus breaks up with emission of a heavy particle (α particle). In such a case the proper disintegration energy would be independent of the energy of the neutron, if the new nucleus were formed in its ground state on all occasions. Kurie puts forward evidence for this result.‡ He assumes that, in all like cases, in which an α particle is emitted, a similar mechanism is involved. Against this assumption is the fact that in two cases at least (Li, B) Fermi[39] has looked for the radiation from any pure capture process and failed to find it. Whilst this does not invalidate the whole point of view, it seems definitely preferable at the moment to admit of no exception to the general rule that when heavy particles are emitted in artificial disintegration the complete synthesis-disintegration process must be regarded energetically as a single event.

* Kurtchatow, Kurtchatow and Latychev[86] have recently confirmed the mode of disintegration of boron by slow neutrons by using the vapour $(CH_3O)_3B$ in an expansion chamber.

† There seems to be some difficulty in accounting for the presence of very fast neutrons in Kurie's experiment.

‡ Kurie's identification of the proper disintegration energy for N^{15} with the energy of the bombarding α particle which enters by resonance into the nucleus B^{11} is invalidated by his omission of a factor 11/15 which should be employed to correct the latter for the motion of the centroid of the system.

By way of summary of the present position, the following facts may be stated. Hitherto, neutrons, protons, deuterons, α particles and lithium ions[90] have been successfully employed as projectiles for disintegration. In one or other of the transformations so produced neutrons, protons, deuterons, H^3 and He^3 nuclei and α particles have been liberated as the lighter (lightest) product nucleus. As heavier products, in the same disintegrations, the following nuclei have appeared: H^3, He^3, He^4, He^5 (?), Li^6, Li^7, etc. It will be noted that these two lists overlap to some extent. With the exception of O^{18} and F^{19}, all the known stable nuclei up to Ne^{22} have been produced in one disintegration process at least, many of them in more than one process. Thus, C^{12} and C^{13} have each been produced in at least three processes by which stable nuclei result. This feature is even more characteristic of the reactions in which unstable species are formed.

§ 6. ARTIFICIAL RADIOACTIVITY

The nature of artificial radioactivity, and the character of some of the transformations in which radioactive species of short life may be artificially produced, have been described in the last section[91], whilst certain of its wider implications were discussed in § 2. Here it remains to give a more systematic—if brief—account of the whole subject, certainly the greatest single addition to the accepted canon of nuclear physics during 1934.

Artificial radioactivity may be produced by bombarding various elements with α particles[92], protons[93], deuterons[94] and neutrons[95]. As far as is at present known, with α particles boron, nitrogen[96], fluorine (?), sodium[97], magnesium, aluminium, silicon[98], phosphorus[97] and potassium[99] yield a radioactive product, with protons boron[100] and carbon, with deuterons lithium[101], boron[102], carbon, nitrogen[103], fluorine[101], sodium[104] and aluminium[105], and with neutrons roughly fifty elements, from fluorine to uranium, may be activated. Radioactivity produced by neutron bombardment will be considered separately, as providing the largest amount of data for discussion; that produced by proton bombardment, since it provides so little, may be summarily disposed of. The reaction with carbon has already been given (reaction (3)), that with boron is undoubtedly similar:

$$ {}_5B^{10} + {}_1H^1 \rightarrow {}_6C^{11} \qquad \text{.....(20).} $$

Both products, C^{11} and N^{13}, emit positrons.

The transformations produced by α particles and deuterons, since they present certain similarities, may be considered together. Many cases are known in which stable species result from such transformations, when protons are emitted. It appears, very frequently, that, with the same elements, the emission of neutrons may occur instead. In such a case the resulting nucleus is usually unstable. It transforms, with emission of a positron, to the stable nucleus known to be produced in the alternative mode of primary disintegration. Thus, alternative to (1), is the transformation

$$ {}_7N^{14} + {}_2He^4 \rightarrow {}_9F^{17} + {}_0n^1 \qquad \text{.....(21),} $$

whilst, with nitrogen and deuterons, either

$$_7N^{14}+{}_1H^2 \rightarrow {}_7N^{15}+{}_1H^1 \quad \text{.....(22),}$$

or

$$_7N^{14}+{}_1H^2 \rightarrow {}_8O^{15}+{}_0n^1 \quad \text{.....(23),}$$

may occur. The nuclei O^{15} and F^{17} transform spontaneously into N^{15} and O^{17}, respectively, with the emission of positrons. Sometimes, when α particles or deuterons are employed, β-active (negative electron-active) substances are produced. This occurs when the bombarded species is such that capture disintegration with the emission of a proton leads to the formation of an unstable nucleus. Thus, Al^{28} and Na^{24} are formed in the reactions

$$_{12}Mg^{25}+{}_2He^4 \rightarrow {}_{13}Al^{28}+{}_1H^1 \quad \text{.....(24),}$$

$$_{11}Na^{23}+{}_1H^2 \rightarrow {}_{11}Na^{24}+{}_1H^1 \quad \text{.....(25),}$$

and subsequently change into Si^{28} and Mg^{24}, respectively, with emission of β particles. The alternative modes of primary disintegration, if they take place, now yield these same stable products and neutrons. For example, instead of (25),

may occur. $$_{11}Na^{23}+{}_1H^2 \rightarrow {}_{12}Mg^{24}+{}_0n^1 \quad \text{.....(26)}$$

So far, without exception, the radioelements formed by neutron bombardment have been found to be β-active. At least three types of primary disintegration have been established: capture of the neutron may be followed by emission of an α particle, a proton or a quantum of γ radiation. In most of the cases in which heavy elements are activated the last named process is effective. This may easily be understood, for, as the atomic number of the nucleus increases, the chance of a heavy charged particle escaping, except in conditions of resonance, rapidly becomes vanishingly small. In practice, the decision between the three possible primary disintegration processes has generally been made by chemical means. Substances which might possibly be isotopic with the active product have been added in small quantities to the irradiated material; after being separated again, they have been individually tested for activity[106]. Now, other methods are available for obtaining a check on these results. The material under test may be irradiated first in air and afterwards with source and target surrounded by a large quantity of some hydrogenous substance, such as water or paraffin. This is most efficient in reducing to quite small values the neutron velocities and it is found that, with simple capture processes, a considerable increase in induced activity results[60]. The actual increase observed in any case depends upon the element and upon the geometrical arrangement employed; it may occasionally be of the order of 100-fold. In these favourable circumstances it is then possible to detect the γ radiation emitted when the neutron is captured, even when the relatively weak neutron sources from Po–Be are employed.*

γ radiation emitted after irradiation was first found with iron, aluminium and silicon by Fleischmann[107]. It has since been shown to be a frequent accompaniment of the β-decay[108]. These artificially produced radioactive substances, there-

* Intense γ rays from the source make this experiment impracticable with the much stronger Ra–Be neutron sources.

fore, behave in general no differently from the β-active bodies of the three main series: often emission of the β particle leaves the nucleus in an excited state, from which it makes transition to the ground state with the emission of radiation. Moreover, the β-active bodies of the main series exhibit another peculiarity of which evidence is not altogether lacking in the new data. In the main series two β-active substances frequently occupy neighbouring positions in the series; one changes directly into the other in the process of disintegration. Evidence that chain products occur amongst the radioactive substances formed from uranium[109] and thorium is perhaps most definite in the latter case[110]. Here, Curie, v. Halban and Preiswerk find that an activity arising in a radioelement of atomic number 91 and half value period about $2\frac{1}{2}$ minutes may be separated, in roughly the same amount, either immediately after irradiation, or when 15 minutes have elapsed. They propose the following scheme:

$$_{90}Th^{232} + {}_{0}n^{1} \rightarrow {}_{90}Th^{233} \qquad \text{.....(27),}$$

$$_{90}Th^{233} \xrightarrow[25\ \text{min.}]{\beta} {}_{91}Pa^{233} \xrightarrow[2{\cdot}5\ \text{min.}]{\beta} {}_{92}U^{233} \longrightarrow \qquad \text{.....(28).}$$

Both radioelements in (28) have mass numbers of the type $4n+1$; it is plausible to assume that they are members of a complete disintegration series, with this specification, of which no trace remains on earth because of the too rapid decay of its ultimate parent, U^{237}. Bombardment of thorium by neutrons results in three activities other than those already described. To explain these, in addition, it is necessary to assume the following transformations:

$$_{90}Th^{232} + {}_{0}n^{1} \rightarrow {}_{88}Ra^{229} + {}_{2}He^{4} \qquad \text{.....(29),}$$

$$_{88}Ra^{229} \xrightarrow{\beta} {}_{89}Ac^{229} \xrightarrow{\beta} {}_{90}Th^{229} \longrightarrow \qquad \text{.....(30)[111],}$$

$$_{90}Th^{232} + {}_{0}n^{1} \rightarrow {}_{89}Ac^{232} + {}_{1}H^{1} \qquad \text{.....(31),}$$

$$_{89}Ac^{232} \xrightarrow{\beta} {}_{90}Th^{232} \longrightarrow \qquad \text{.....(32).}$$

The chain products of (30), likewise of the type $4n+1$, may naturally result from an $\alpha\beta$ branching at Th^{233} in the hypothetical series above discussed. As concerns the primary transformations (29), (31) it may be noted that, in these, heavy particles are emitted. It is too early yet to know whether conditions of resonance are involved. The artificial radioactivity obtained with uranium is especially interesting from another point of view. As before, at least four periods are required in an analysis of the observed decay of activity. There is reason to believe that all but one of these are to be ascribed to hitherto unknown trans-uranic elements ($Z > 92$). This point of view has been criticized by Grosse and Agruss[112], but their criticism has since been withdrawn[113]. In this connexion Grosse[114] has suggested that the inconclusive nature of certain chemical tests may be due to the fact that a new series of "rare earth" elements begins with uranium ($Z = 92$).

In a final review of the subject of artificial radioactivity attention may again be directed to the many cases in which the same unstable nucleus may be produced in a number of different primary disintegration processes. Al^{28}, for example, may be obtained in five different ways—by bombarding magnesium with α particles (24),

aluminium with deuterons or neutrons and silicon or phosphorus with neutrons. So far there is no evidence outstanding* which indicates any difference in the radioactive properties in any such instance, when different methods of production are employed. It seems that the two numbers A and Z specify uniquely the properties of an atomic nucleus. If nuclear isomerism[116] is to be found in nature, its incidence must be extremely rare indeed[117].

§ 7. NUCLEAR TRANSFORMATION BY γ RADIATION

The first conclusive evidence of specific interaction between quanta and atomic nuclei was obtained in 1934. Previously many possibilities of interaction had been discussed but those which were found were always of a general nature. The "extra" absorption of hard γ rays, for example, varies regularly from element to element[118]. It may be styled "nuclear" absorption only because the neighbourhood of an atomic nucleus is the one place where electric fields are to be found sufficiently intense for the process to occur with reasonable probability[119]. But it does not appear that the nucleus is in any way affected in the act of absorption. Recently, however, two cases of specific interaction have been recorded. Chadwick and Goldhaber[120] showed that protons are liberated when heavy hydrogen gas is subjected to the action of γ rays of more than about 2×10^6 eV. energy. This, then, is the energy balance (Q) entering into the nuclear equation

$$ {}_1H^2 + Q = {}_1H^1 + {}_0n^1 \qquad \text{.....(33);} $$

its exact determination provides the best method of obtaining the mass of the neutron in terms of the masses of the hydrogen isotopes, regarded as known. The second case of photo-disintegration, that of beryllium, was investigated first by Szilard and Chalmers[121]. These authors detected the neutrons which were emitted. So far, however, it is not entirely certain whether the charged residue in this process is the nucleus Be^8 or two α particles.

Photo-disintegration with the emission of neutrons is the direct reverse of transformation by radiative capture, which has already been discussed. When sufficient knowledge is available concerning the particles involved, the probability of each process may be separately evaluated. This has, until now, been done only for the simplest case, (33)[122]. On the other hand, relative probabilities may be calculated from thermodynamical reasoning alone, if the energy balance is known[120]. No doubt when γ-ray sources of high quantum energy and sufficient intensity become available many other examples of photo-disintegration will be established; with natural radioactive sources only, however, it appears that the possibilities, amongst the light elements at least, are already exhausted in H^2 and Be^9. Nevertheless, quanta of 10^7 eV. energy and upwards are known to be present as one component of the penetrating radiation at sea level[123]; it is possible that evidence for photo-disintegration produced by this radiation will not long continue to be lacking.

* The discrepancy which appeared to exist between the periods of N^{13} formed by the bombardment of boron by α particles and of carbon by protons (3) and deuterons has been shown to be illusory[115].

REFERENCES

(1) ANDRADE. *Reports on Progress in Physics*, **1**, 269 (1934). Hereinafter referred to as *Report* (1933).
(2) NODDACK. *Angew. Chem.* **47**, 301 (1934).
(3) LIBBY. *Phys. Rev.* **46**, 196 (1934).
(4) GUEBEN. Thèse, Liége (1933).
(5) DEMPSTER. *Nature*, Lond., **135**, 993; **136**, 65 (1935).
(6) ASTON. *Proc. roy. Soc.* **149**, 396 (1935).
(7) KIRCHNER. *Naturwissenschaften*, **22**, 480 (1934).
CRANE, DELSASSO, FOWLER and LAURITSEN. *Phys. Rev.* **47**, 971 (1935).
BONNER and BRUBAKER. *Phys. Rev.* **47**, 973 (1935).
OLIPHANT, KEMPTON and RUTHERFORD. *Proc. roy. Soc.* **150**, 241 (1935).
(8) *Report* (1933), p. 314.
(9) HEVESY. *Nature*, Lond., **135**, 96 (1935).
KLEMPERER. *Proc. roy. Soc.* **148**, 638 (1935).
(10) SCHÜLER and SCHMIDT. *Z. Phys.* **92**, 148 (1934).
(11) *Report* (1933), p. 281.
SCHÜLER and GOLLNOW. *Naturwissenschaften*, **22**, 511 (1934).
GROSSE. *Proc. roy. Soc.* **150**, 363 (1935).
(12) *Report* (1933), p. 271.
(13) OLIPHANT, KEMPTON and RUTHERFORD. *Proc. roy. Soc.* **150**, 241 (1935).
(14) BETHE. *Phys. Rev.* **47**, 633, 795 (1935).
(15) ASTON. *Nature*, Lond., **135**, 541 (1935).
(16) *Report* (1933), p. 283.
(17) *Report* (1933), p. 274 and § 9.
(18) Cf. RACAH. *Z. Phys.* **71**, 431 (1931).
(19) CASIMIR. *Physica*, 's Grav., **2**, 719 (1935).
(20) SCHÜLER and SCHMIDT. *Z. Phys.* **94**, 457; **95**, 265 (1935).
(21) ASTON. *Isotopes*, Chap. XI (1923).
(22) HERTZ. *Naturwissenschaften*, **21**, 884 (1933).
(23) HARMSEN, HERTZ and SCHÜTZE. *Z. Phys.* **90**, 703 (1934).
(24) *Report* (1933), p. 284.
FARKAS. *Light and Heavy Hydrogen*, Part II, Chap. II (1935).
(25) HERTZ. *Z. Phys.* **91**, 810 (1934).
(26) SMYTHE, RUMBAUGH and WEST. *Phys. Rev.* **45**, 724 (1934).
(27) OLIPHANT, SHIRE and CROWTHER. *Proc. roy. Soc.* **146**, 922 (1934).
(28) SZILARD and CHALMERS. *Nature*, Lond., **134**, 462 (1934).
(29) AMALDI and others. *Proc. roy. Soc.* **149**, 522 (1935).
(30) *Report* (1933), § 9.
(31) FRISCH and STERN. *Z. Phys.* **85**, 4 (1933).
ESTERMANN and STERN. *Z. Phys.* **85**, 17; **86**, 132 (1933).
—— —— *Nature*, Lond., **133**, 911 (1934).
(32) BREIT and RABI. *Phys. Rev.* **38**, 2082 (1931).
(33) COHEN. *Phys. Rev.* **46**, 713 (1934).
(34) MILLMAN. *Phys. Rev.* **47**, 739 (1935).
(35) RABI, KELLOGG and ZACHARIAS. *Phys. Rev.* **46**, 157, 163 (1934).
(36) MILLMAN, FOX and RABI. *Phys. Rev.* **46**, 320 (1934).
(37) RABI and COHEN. *Phys. Rev.* **46**, 707 (1934).
(38) RABI. *Phys. Rev.* **47**, 338 (1935).
(39) BACK and GOUDSMIT. *Z. Phys.* **47**, 174 (1928).
(40) HEYDENBURG, LARRICK and ELLETT. *Phys. Rev.* **40**, 1041 (1932).
(41) BREIT. *Rev. mod. Phys.* **5**, 91 (VII, § 4) (1933).
(42) LARRICK. *Phys. Rev.* **46**, 581 (1934).
ELLETT and HEYDENBURG. *Phys. Rev.* **46**, 583 (1934).
(43) HEYDENBURG. *Phys. Rev.* **46**, 802 (1934).

(44) KALCKAR and TELLER. *Nature*, Lond., **134**, 180 (1934).
—— —— *Proc. roy. Soc.* **150**, 520 (1935).
(45) FARKAS, FARKAS and HARTECK. *Proc. roy. Soc.* **144**, 481 (1934).
(46) FARKAS and FARKAS. *Nature*, Lond., **135**, 372 (1935).
(47) BREIT. *Phys. Rev.* **46**, 319 (1934).
(48) VENKATESACHAR and SIBAIYA. *Nature*, Lond., **136**, 65 (1935).
FUCHS and KOPFERMANN. *Naturwissenschaften*, **23**, 372 (1935).
(49) *Report* (1933), §§ 5, 6, 8.
(50) *Report* (1933), § 8.
(51) SAVEL. *Ann. Phys.*, Paris, **4**, 88 (1935).
(52) SCHNETZLER. *Z. Phys.* **95**, 302 (1935).
(53) SAVEL. *C.R. Acad. Sci.*, Paris, **198**, 1404 (1934).
(54) —— *C.R. Acad. Sci.*, Paris, **196**, 1482 (1933).
(55) —— *C.R. Acad. Sci.*, Paris, **198**, 368 (1934).
(56) *Report* (1933), p. 294.
(57) Cf. CHADWICK and FEATHER. *Int. Conf. Phys.* **1**, 95 (1935).
LEA. *Proc. roy. Soc.* **150**, 637 (1935).
(58) DUNNING. *Phys. Rev.* **45**, 586 (1934).
(59) BECK. *Phys. Z.* **31**, 945 (1930).
WILSON. *Mon. Not. R. astr. Soc.* **91**, 283 (1931).
MASSEY and MOHR. *Nature*, Lond., **133**, 211 (1934).
—— —— *Proc. roy. Soc.* **148**, 206 (1935).
(60) FERMI, and others. *La Ricerca Scientifica*, **5** (2), 282 (1934).
(61) COCKCROFT, GILBERT and WALTON. *Nature*, Lond., **133**, 328 (1934).
—— —— —— *Proc. roy. Soc.* **148**, 225 (1935).
(62) CRANE, DELSASSO, FOWLER and LAURITSEN. *Phys. Rev.* **47**, 782 (1935).
—— —— —— —— *Phys. Rev.* **48**, 102 (1935).
(63) —— —— —— —— *Phys. Rev.* **46**, 531 (1934).
(64) CRANE and LAURITSEN. *Int. Conf. Phys.* **1**, 130 (1935).
(65) HAFSTAD and TUVE. *Phys. Rev.* **47**, 506 (1935).
CRANE and LAURITSEN. *Phys. Rev.* **47**, 420 (1935).
(66) BREIT and YOST. *Phys. Rev.* **46**, 1110 (1934).
—— —— *Phys. Rev.* **47**, 508 (1935).
(67) BETHE. *Phys. Rev.* **47**, 747 (1935).
(68) OPPENHEIMER. *Phys. Rev.* **47**, 845 (1935).
(69) FEATHER. *Proc. roy. Soc.* **136**, 709 (1932).
FERMI and others. *Proc. roy. Soc.* **146**, 483 (1934).
(70) CHADWICK, FEATHER and DAVIES. *Proc. Camb. phil. Soc.* **30**, 357 (1934).
(71) *Report* (1933), § 11.
(72) Cf. FEATHER. *Ann. Rep. Chem. Soc.* **31**, 368 (1935).
(73) CHADWICK and FEATHER. *Int. Conf. Phys.* **1**, 95 (1935).
(74) FEATHER. *Proc. roy. Soc.* **142**, 689 (1933).
(75) OLIPHANT. *Int. Conf. Phys.* **1**, 144 (1935).
OLIPHANT, KEMPTON and RUTHERFORD. *Proc. roy. Soc.* **149**, 406 (1935).
See also references (13) and (27).
(76) COCKCROFT. *Int. Conf. Phys.* **1**, 112 (1935).
(77) DEE and WALTON. *Proc. roy. Soc.* **141**, 733 (1933).
DEE. *Proc. roy. Soc.* **148**, 623 (1935).
DEE and GILBERT. *Proc. roy. Soc.* **149**, 200 (1935).
(78) HENDERSON, LIVINGSTON and LAWRENCE. *Phys. Rev.* **46**, 38 (1934).
MCMILLAN. *Phys. Rev.* **46**, 325, 868 (1934).
LAWRENCE. *Phys. Rev.* **47**, 17 (1935).
LAWRENCE, MCMILLAN and HENDERSON. *Phys. Rev.* **47**, 273 (1935).
MCMILLAN and LAWRENCE. *Phys. Rev.* **47**, 343 (1935).
(79) CRANE, DELSASSO, FOWLER and LAURITSEN. *Phys. Rev.* **46**, 1109 (1934).
—— —— —— —— *Phys. Rev.* **48**, 100 (1935).
See also references (64), (62) and (63).

(80) Matzner. *S.B. Akad. Wiss. Wien*, **143**, 579 (1934).
(81) Chadwick and Goldhaber. *Nature*, Lond., **135**, 65 (1935).
(82) Amaldi. *Nuovo Cim.* **12**, 223 (1935).
(83) Taylor and Goldhaber. *Nature*, Lond., **135**, 341 (1935).
(84) Blau and Wambacher. *S.B. Akad. Wiss. Wien*, **143**, 401 (1934).
Blau. *J. Phys. Radium*, **5**, 61 (1934).
(85) Taylor. *Proc. roy. Soc.* **150**, 382 (1935).
(86) Kurtchatow, Kurtchatow and Latychev. *C.R. Acad. Sci.*, Paris, **200**, 1199 (1935).
(87) Harkins, Gans and Newson. *Phys. Rev.* **47**, 52 (1935).
(88) Kurie. *Phys. Rev.* **45**, 904 (1934).
(89) —— *Phys. Rev.* **47**, 97 (1935).
(90) Zeleny, Brasefield, Bock and Pollard. *Phys. Rev.* **46**, 318 (1934).
Whitmer and Pool. *Phys. Rev.* **47**, 795 (1935).
(91) Cf. Fermi. *Int. Conf. Phys.* **1**, 75 (1935).
Curie-Joliot and Joliot. *Int. Conf. Phys.* **1**, 78 (1935).
(92) Curie and Joliot. *C.R. Acad. Sci.*, Paris, **198**, 254 (1934).
(93) Cockcroft, Gilbert and Walton. *Nature*, Lond., **133**, 328 (1934).
(94) Lauritsen, Crane and Harper. *Science*, **79**, 234 (1934).
(95) Fermi. *La Ricerca Scientifica*, **5** (1), 283 (1934).
(96) Wertenstein. *Nature*, Lond., **133**, 564 (1934).
(97) Frisch. *Nature*, Lond., **133**, 721 (1934).
(98) Fahlenbrach. *Naturwissenschaften*, **23**, 288 (1935).
(99) Żyw. *Nature*, Lond., **134**, 64 (1934).
(100) Crane, Delsasso, Fowler and Lauritsen. *Phys. Rev.* **48**, 102 (1935).
(101) —— —— —— —— *Phys. Rev.* **47**, 971 (1935).
(102) —— —— —— —— *Phys. Rev.* **47**, 887 (1935).
(103) Livingston and McMillan. *Phys. Rev.* **46**, 437 (1934).
McMillan and Livingston. *Phys. Rev.* **47**, 452 (1935).
(104) Lawrence. *Phys. Rev.* **47**, 17 (1935).
(105) McMillan and Lawrence. *Phys. Rev.* **47**, 343 (1935).
(106) Fermi and others. *Proc. roy. Soc.* **146**, 483 (1934).
(107) Fleischmann. *Naturwissenschaften*, **22**, 434 (1934).
(108) Amaldi and others. *La Ricerca Scientifica*, **5** (1), 452 (1934).
(109) Fermi. *Nature*, Lond., **133**, 898 (1934).
(110) Curie, v. Halban and Preiswerk. *C.R. Acad. Sci.*, Paris, **200**, 1841 (1935).
—— —— —— *C.R. Acad. Sci.*, Paris, **200**, 2079 (1935).
(111) Hahn and Meitner. *Naturwissenschaften*, **23**, 320 (1935).
(112) Grosse and Agruss. *Nature*, Lond., **134**, 773 (1934).
(113) —— —— *J. Amer. chem. Soc.* **57**, 438 (1935).
(114) Grosse. *J. Amer. chem. Soc.* **57**, 440 (1935).
(115) Ellis and Henderson. *Nature*, Lond., **135**, 429 (1935).
Fahlenbrach. *Z. Phys.* **94**, 607 (1935).
(116) Gamow. *Int. Conf. Phys.* **1**, 60 (1935).
(117) Cf. Sosnowski. *C.R. Acad. Sci.*, Paris, **200**, 1027 (1935).
(118) Tarrant. *Proc. roy. Soc.* **135**, 223 (1932).
Jacobsen. *Z. Phys.* **70**, 145 (1931).
McMillan. *Phys. Rev.* **46**, 325, 868 (1934).
(119) Bethe and Heitler. *Proc. roy. Soc.* **146**, 83 (1934).
(120) Chadwick and Goldhaber. *Nature*, Lond., **134**, 237 (1934).
(121) Szilard and Chalmers. *Nature*, Lond., **134**, 494 (1934).
(122) Bethe and Peierls. *Proc. roy. Soc.* **148**, 146 (1935).
Massey and Mohr. *Proc. roy. Soc.* **148**, 206 (1935).
(123) Anderson, Millikan, Neddermeyer and Pickering. *Phys. Rev.* **45**, 352 (1934).

GEOPHYSICAL PROSPECTING

By E. LANCASTER-JONES, B.A.

§ 1. INTRODUCTION

THEORETICALLY, the expression "geophysical prospecting" is used to distinguish those methods of determining the location of valuable mineral deposits by means of measurements of physical magnitudes which are modified from a known standard by the nature and position of the deposit. Practically, it is understood that methods involving actual vision of, or contact with, the deposit are excluded, although such methods are obviously "physical". Similarly the ordinary physical measurements of geological exploration are excluded, and the term "geophysical" confined to those magnitudes, namely gravitational, electrical, magnetic and seismic, which can be detected and measured at a distance and in spite of the intervention of the opaque covering of soil and rock which normally precludes actual vision or effective contact with the deposit.

It is understood also that the magnitudes above specified may be modified abnormally not only by the deposit itself, but also by configurations of strata closely linked with the deposit; and indeed the influence of the deposit may be quite subordinate or even negligible, and yet the existence and locality of the deposit may be determined indirectly by geophysical measurements of the surrounding, associated strata.

The subject therefore covers a wide field of physical phenomena and measurement of the physical characteristics of minerals and associated rocks in their natural, highly complicated states. Except for comparative purposes, we are not concerned with laboratory measurements of carefully isolated and purified specimens, but with field determinations of the aggregated effect of the natural, complex domain in which mineralization is suspected.

The peculiar difficulties of geophysical prospecting consist in the minuteness of the magnitudes measured under the exigencies of field work, and the fact that the effects are aggregated; it is impossible to isolate the deposit sought from its surroundings. Highly sensitive apparatus and elaborate technique are therefore essential to geophysical prospecting, and in the need for this refinement lies the explanation why the development of the methods has taken place almost entirely within the present century, and indeed within the last fifteen years.

It is true that crude attempts were made in the nineteenth century to employ both electrical and magnetic methods of prospecting, but the apparatus was quite inadequate except in a few cases of pronounced anomalies which could hardly escape detection even with the insensitive instruments then available. These early efforts may be ignored, since they had little relation to the real development of the subject.

Similarly, methods involving the personal characteristics of the observer, such as those using the divining-rod and its like, may be dismissed as irrelevant. Whatever value lies in divination, these methods cannot be correlated with any known physical magnitude independently measurable by any suitably trained physicist.

The present article will therefore be restricted to the recent, fundamental development of those gravitational, magnetic, electrical and seismic methods which represent the real field of geophysical prospecting.

Only a brief digest of the development can be given here. Already geophysical prospecting has its text-books, the more important of which are given in the references at the end of this article. Even in these text-books of hundreds of pages, full justice cannot be done to the researches and investigations, theoretical, in the laboratory and in the field, many of which have not been published, that constitute the complete sum of our modern knowledge of a subject which owes much to the past and to the general total of physical knowledge, but which has evolved a technique and apparatus of its own. In many respects the apparatus of the applied geophysicist, combining as it does extreme sensitivity with astonishing portability and field-worthiness, calls for the utmost skill of the designer and the instrument maker, and embodies lessons worthy of closer study by physicists in general.

§ 2. THE GRAVITATIONAL METHOD

The force of gravity, due to the attraction of the whole earth, etc., varies both in magnitude and direction according to the position of the point at which it is measured. This fact has been familiar to physicists for very many years. The principal variations have been measured by means of the gravity pendulum for the magnitude, and the spirit level for the direction, at a great many places on the earth's surface. The variations so measured can be attributed partly to the deviation of the earth from perfect sphericity, to its rotation about its axis, to the presence of large obvious upheavals in mountainous regions and depressions in oceanic regions, and to the residual anomalies of density below the earth's surface.

The pendulum and spirit level are only capable of detecting variations in gravity due to very extensive anomalies of crustal density. It was not until 1888 that the genius of Eötvös evolved an instrument, the gravity torsion balance, which could detect the very small variations in the force of gravity of a local, regional character, which arose from very local anomalies of density such as those associated with mineral deposits at depths below the surface of only a few hundred feet.

The essence of the discovery of Eötvös lies not so much in his modification of the well-known torsion balance of Coulomb, as in his realization that local anomalies of density affected to a measurable extent the very functions of the gravity force which operated upon a torsion balance, which hitherto had been used solely to measure forces originating in the direct horizontal pull on one of the balanced masses due to a second mass placed adjacent to the former. Eötvös, on the other hand, employed the torsion balance to measure the differential pull exerted by the whole earth upon the two balanced masses, in virtue of their slightly different position with respect to the earth. The principle is readily seen if we recognize that

the earth gives rise to a field of force, gravity, which varies uniformly in the small neighbourhood surrounding the balance. The field has a potential function U.

The Eötvös balance. First or Coulomb type. Referring to figure 1, O is the centre of gravity of the beam of a simple Coulomb balance, which has balanced masses m, m' at its extremities A, A', and is suspended at O by a vertical torsion wire attached to an adjustable torsion head. Select rectangular axes Ox along OA, Oy the other horizontal axis and Oz vertically downwards.

Since U is the potential, we have at O

$$\left(\frac{dU}{dx}\right)_0 = O; \quad \left(\frac{dU}{dy}\right)_0 = O; \quad \left(\frac{dU}{dz}\right)_0 = g_0,$$

where g_0 is the force of gravity at O.

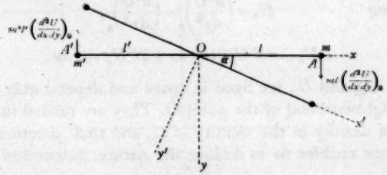


Figure 1.

At A, where $OA = l$,

$$\left(\frac{dU}{dx}\right)_A = \left(\frac{dU}{dx}\right)_0 + l\left(\frac{d}{dx}\frac{dU}{dx}\right)_0 = l\left(\frac{d^2U}{dx^2}\right)_0 \text{ acting along } OA,$$

$$\left(\frac{dU}{dy}\right)_A = \left(\frac{dU}{dy}\right)_0 + l\left(\frac{d}{dx}\frac{dU}{dy}\right)_0 = l\left(\frac{d^2U}{dx\,dy}\right)_0 \text{ acting } \perp OA \text{ in the sense shown.}$$

At A', where $OA' = -l'$, similarly

$$\left(\frac{dU}{dx}\right)_{A'} = \left(\frac{dU}{dx}\right)_0 - l'\left(\frac{d}{dx}\frac{dU}{dx}\right)_0 = l'\left(\frac{d^2U}{dx^2}\right)_0 \text{ acting along } OA',$$

$$\left(\frac{dU}{dy}\right)_{A'} = \left(\frac{dU}{dy}\right)_0 - l'\left(\frac{d}{dx}\frac{dU}{dy}\right)_0 = l'\left(\frac{d^2U}{dx\,dy}\right)_0 \text{ acting } \perp OA' \text{ in the sense shown.}$$

There is thus a twisting torque due to the two forces $ml\left(\frac{d^2U}{dx\,dy}\right)_0$, $m'l'\left(\frac{d^2U}{dx\,dy}\right)_0$ tending to rotate the arm $A'OA$ about the suspension axis Oz.

Putting U_{xy} for $\left(\frac{d^2U}{dx\,dy}\right)_0$, the torque is $(ml^2 + m'l'^2)\,U_{xy} = I'U_{xy}$, where I' is the moment of inertia of the beam system about the axis Oy. This must not be confused with the moment of inertia I about Oz, although, for the above system, in which the masses and beam are concentrated along a line, $I' = I$.

Now, let the axes Ox, Oy remain fixed in space, and axes Ox', Oy' move with the beam $A'OA$ when it is rotated into a new azimuth (figure 1) wherein Ox' makes an angle α with Ox.

We have the torque now $I'U_{x'y'}$ and, for converting $U_{x'y'}$ into differentials relating to the system (Ox, Oy), we have the standard relations

$$\frac{d}{dx'} = \cos\alpha \frac{d}{dx} + \sin\alpha \frac{d}{dy},$$

$$\frac{d}{dy'} = -\sin\alpha \frac{d}{dx} + \cos\alpha \frac{d}{dy},$$

whence

$$\frac{d}{dx'}\frac{d}{dy'} = \cos\alpha \frac{d}{dx}\left(-\sin\alpha \frac{d}{dx} + \cos\alpha \frac{d}{dy}\right) + \sin\alpha \frac{d}{dy}\left(-\sin\alpha \frac{d}{dx} + \cos\alpha \frac{d}{dy}\right),$$

therefore

$$\frac{d^2}{dx'dy'} = \cos 2\alpha \frac{d^2}{dxdy} + \frac{1}{2}\sin 2\alpha\left(\frac{d^2}{dy^2} - \frac{d^2}{dx^2}\right).$$

Therefore, putting

$$U_\Delta = \left(\frac{d^2U}{dy^2}\right)_0 - \left(\frac{d^2U}{dx^2}\right)_0,$$

we get

$$I'U_{x'y'} = I'U_{xy}\cos 2\alpha + \tfrac{1}{2}I'U_\Delta \sin 2\alpha.$$

The magnitudes U_{xy} and U_Δ are fixed in space and depend only upon the field of gravity in the neighbourhood of the point O. They are related intimately with the local anomalies of density in the vicinity of O, and their determination at a series of different stations enables us to deduce the nature, extent and position of such anomalies.

I' is an instrumental constant, determined in the laboratory. The effect of the torque $I'U_{x'y'}$ is balanced by a twist of the torsion wire through a very small angle, measured by a deflection of a telescope cross-wire upon a scale which is fixed to the balance. If we denote the reading of the scale in the azimuth α by the magnitude n_α, and the reading when the wire is untwisted by n, then the deflection in radians is $(n_\alpha - n)/2D$, where D is the distance, measured in scale units, of the scale from the mirror which is fixed to the beam in the axis of the torsion wire. If the torsional rigidity of the wire is τ, the restoring torque is measured by $\tau(n_\alpha - n)/2D$ and this must equal the disturbing torque $I'U_{x'y'}$. Whence

$$n_\alpha - n = \frac{2DI'}{\tau} U_{xy}\cos 2\alpha + \frac{2DI'}{\tau}\frac{U_\Delta}{2}\sin 2\alpha,$$

which represents the fundamental equation for the Eötvös balance of the first type, i.e. the Coulomb type.

Here the quantity $2DI'/\tau$ is an instrumental constant. To determine U_{xy} and U_Δ at the station it is only necessary to set the balance in a series of azimuth settings α and determine the equilibrium reading n_α at each setting.

It may be noted that n is to some extent slowly variable with time, so that there are really three unknowns n, U_{xy} and U_Δ which need determining at each station. The usual procedure is to set the beam in successive azimuths $\alpha = 0°$, $\alpha = 120°$ and $\alpha = 240°$, and to determine n as the mean of the readings n_α in these three successive azimuths. This value of n is taken as valid for the central azimuth occupied in the run, a correct assumption if n varies uniformly with time.

The Eötvös balance. Second type. The type of torsional balance generally known as the Eötvös Balance is the one now to be described. Schematically it differs from the first or Coulomb type in having one of its masses m suspended below the end of the beam A to a depth h. Figure 2 shows the vertical section of this balance.

The force now acting on unit mass m in the direction of the axis Oy, and therefore tending to twist the beam, is

$$l\left(\frac{d^2U}{dx\,dy}\right)_0+h\frac{d}{dz}\left(\frac{dU}{dy}\right)_0,$$

since there is an extra component due to the depth h of m below the plane xOy. The total force Oy on m is thus

$$mlU_{xy}+mhU_{yz}, \text{ where } U_{yz}=\frac{d}{dz}\left(\frac{dU}{dy}\right)_0=\left(\frac{d^2U}{dy\,dz}\right)_0.$$

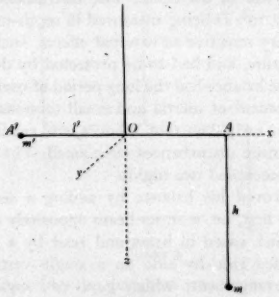


Figure 2.

The total torque about Oz becomes

$$I'U_{xy}+mhlU_{yz}.$$

When the beam is rotated as a whole into a new azimuth Ox', making an angle α with Ox as before, we have $U_{x'y'}$ for U_{xy} and $U_{y'z}$ for U_{yz} in the above, and since Oz remains common to both systems of axes

$$U_{y'z}=\frac{d}{dy'}\frac{dU}{dz}=-\sin\alpha\frac{d}{dx}\frac{dU}{dz}+\cos\alpha\frac{d}{dy}\frac{dU}{dz}$$
$$=-\sin\alpha\,U_{xz}+\cos\alpha U_{yz}.$$

$U_{x'y'}$ is transformed as before, and the total torque becomes

$$I'U_{xy}\cos 2\alpha+\tfrac{1}{2}I'U_{\Delta}\sin 2\alpha-mhlU_{xz}\sin\alpha+mhlU_{yz}\cos\alpha.$$

The equation of equilibrium of the system is now

$$n_\alpha-n=\frac{2DI'}{\tau}U_{xy}\cos 2\alpha+\frac{2DI'}{\tau}\frac{U_\Delta}{2}\sin 2\alpha-\frac{2Dmhl}{\tau}U_{xz}\sin\alpha+\frac{2Dmhl}{\tau}U_{yz}\cos\alpha$$

and there are 5 unknowns, n, U_{xy}, U_Δ, U_{xz} and U_{yz}, which can be determined from 5 settings of the balance.

A single-beam balance of this type was the one first used by Eötvös for measuring local variations in the field of gravity in certain districts where geological structures gave rise to gravitational anomalies. A celebrated instance was his measurements in the Arad district of Transylvania.

Eötvös showed by computation that the magnitudes U_{zy}, etc. were normally of the order 10^{-9} c.g.s. unit, as compared with ordinary values of g (i.e. dU/dz) which are of the order 10^{3} units. His balance had therefore to be very sensitive to detect such small quantities, and he obtained the requisite sensitivity by the use of a very fine platinum-iridium torsion wire, which could easily sustain the total load of the beam, about 80 gm., and yet had a torsional rigidity constant τ of only about 0·4 unit. He used a scale divided into millimetres and read to tenth-millimetres by means of a telescope; the scale was fixed at a distance 60 cm. from the mirror attached to the axis of the beam. The instrumental constants $2DI'/\tau$ and $2Dmhl/\tau$ were about 2×10^{9}, D being measured in tenth-millimetres.

The system was very sensitive to external effects, such as air-currents caused by changes of temperature, and had to be protected by double, and subsequently triple brass casings. The balance had the long period of oscillation of about 30 min., owing to the large moment of inertia and small torsional rigidity. Each setting of the instrument in the early days took 2 hours, and readings could only be taken at night, when temperature disturbances were small. The complete determination at each station usually occupied two nights.

Later, Eötvös improved his balance by adding a second independent unit, oriented at 180° to the first, i.e. with its beam oppositely directed. Each unit was independently suspended, cased in brass and read by a separate telescope. The two units were mounted side by side on a single vertical axis upon a single column. With this arrangement, which gave two equations at each setting, the number of settings needed was almost halved; not quite, because a new unknown was introduced by the second torsion wire. The quantities U_{zy}, U_{Δ}, etc. are the same for both units, since the two centres of the units are so close together that variations in the gravity field in their neighbourhood are insignificant.

Eötvös thus arrived at the stage when he could determine one station each night.

Schweydar torsion balance, *c.* 1920. The next marked advance in instrumental design was made by Schweydar and the firm of Bamberg, who produced a double-beam, automatic-rotating, photographic-registering instrument. The damping of each balance system was efficient, and the two units were arranged to rotate automatically into their correct azimuth settings every hour. At the end of the hour the readings of two luminous spots, one for each unit, were recorded photographically on a single plate. When seven such readings had been made, the determination was completed. The lateral deviations in the positions of the spots at various settings were measured by means of a graticule, and the determination of U_{zy}, etc. followed simply. The Bamberg instrument could be operated in the daytime as well as at night, and two stations daily became a regular programme.

Schweydar-Askania Z-beam balance, 1926. A further ingenious improvement devised by Schweydar and executed by Bamberg consisted in the introduction of

a rigid beam, with projecting vertical upper and lower rods supporting the two masses m, m' (figure 3), to replace the ordinary beam and lower weight system of Eötvös. The effect is the same, but the arrangement is much more compact, the vertical dimensions of the whole balance being considerably reduced.

Oertling torsion balance, 1925. This was the first torsion balance of English manufacture and possesses several interesting features. It is automatic-rotating, and either visual- or photographic-recording, at the will of the observer. It has an optical system which projects the luminous image of a scale and cross-wire on to a screen or plate with a clarity sufficient for easy reading at a distance of 2 or 3 ft. The instrument is housed within an aluminium tent of ingenious, compact design, to which it can be clamped on to a special trolley for transport. The beam system is of the normal Eötvös type, and the instrument in this feature conforms to the Süss type, whilst in its optical arrangements it is more similar to the Bamberg model.

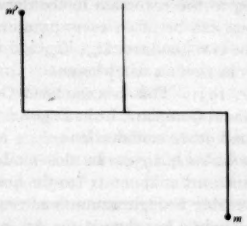


Figure 3.

A special feature introduced by the designers was the use of strip instead of round torsion wires. Greater sensitivity was obtained with greater load-bearing capacity by this modification. The benefit of these wires was, however, better revealed in the gradiometer described below.

Rybar-Süss balance, 1929. Whilst the firm of Bamberg had pinned its faith to photographic, automatic balances, the firm of Süss, who had made the original balances for Eötvös, remained faithful to his visual-recording, non-automatic arrangements, and only modified his double-beam model in minor directions, principally in introducing aluminium casings for brass, as Bamberg had done. Recently, however, they have produced a model designed by Rybar, which is very compact, photographic and automatic. Rybar obtains his compactness by a general reduction in dimensions compensated to preserve the sensitivity by the use of a specially sensitive torsion wire and an optical system of increased power. The beam system remains as before, with one mass fixed to the beam and the other suspended below it.

Oertling gradiometer, 1927. In this instrument the suspended mass system is so arranged that the balance is insensitive to the gravitational magnitudes U_{zz} and U_{Δ}. This arises because the instrumental constant corresponding to I' vanishes. When the masses comprising the suspended system are not concentrated in the

plane $x'Oz$, but are distributed symmetrically on either side of this plane, the constant in question becomes $\int dm\,(y'^2-x'^2)$, where (x', y', z') are the co-ordinates of the element of mass dm and the integration is taken throughout the suspended mass system. It is easy to discover a number of arrangements which satisfy the condition that this integral shall vanish, e.g. a series of equal masses located at the vertices of a regular horizontal polygon of three or more sides, centre at O. The instrument then becomes sensitive only to the "gradients" of gravity (see later) U_{xz} and U_{yz}. The Oertling gradiometer employs a single beam, of circular shape, with two masses on the beam and a third mass supported by means of a vertical tubular aluminium rod at a height roughly 40 cm. above the beam. The radial arm l is here only 4·5 cm. instead of the usual Eötvös 20 cm. Sensitivity is obtained by using a flat strip of wire for which $\tau = 0{\cdot}08$ approximately, and the resulting balance is very compact, robust and speedy in operation. The period of oscillation is much smaller than usual, owing to the reduction in the dimension l, and, with strong damping, azimuth readings can be taken every 20 min. Only three azimuths are necessary to determine the two gradients U_{xz}, U_{yz} and the wire zero n. A station can be observed normally in about 2 to $2\frac{1}{2}$ hours.

Cambridge gradiometer, 1931. This is a modified Oertling gradiometer, made by the Cambridge Instrument Company, Ltd., England. It has an improved, more sensitive, optical system and other modifications.

Askania inclined-beam balance, 1934. In this model, the Z-shaped beam is replaced by a single rod inclined at about 45° to the horizontal, and the period of oscillation is reduced to enable a single azimuth setting to be observed in about 20 min. The sensitivity is slightly less than in the Askania Z-beam model.

In addition to the above instruments, many other models have been devised and constructed, with varying success, the objective being generally to reduce the overall dimensions and speed up the operation of determining the gravity magnitudes at a station. Several projects for assembling a series of beams, symmetrically oriented, on a single pedestal have revealed more fertility in imagination than knowledge of practical construction on the part of their proposers.

With modern instruments a station can be observed in a few hours, and, bearing in mind the time necessary for computation, subsidiary topographical measurements and transport, there is little need for further progress in the direction of reducing the time of observation.

There are, however, many problems interesting to physicists involved in the gravity torsion balance. The great sensitivity of the wires employed introduces disturbances from effects which are ordinarily inactive in physical instruments similarly employing torsion wires. Of these effects, the most important arises from even quite small variations of temperature, which set up eddy currents of which the strength appears to depend on the temperature gradient inside the beam-chamber. Small impurities in the metals or non-metals employed in the construction of the beam may give rise to grave troubles, usually magnetic in character. Rankine has shown that electrostatic effects may cause trouble where mica is employed as a damping device.

Computation of observed gravity values. The reduction of the readings obtained in the instrument to provide the desired values of U_{xy}, etc. is a simple matter of substitution in standard formulae derived from the fundamental equation by the substitution of numerical constants for the instrumental constants and the normal series of values of the azimuth angle α.

Reduction of observed values. The values of the gravity magnitudes as obtained relate to the total gravitational field of the earth, and include the effects of its rotation. From these must be subtracted the so-called "normal values", which depend on the latitude of the station and are given by standard formulae. Subsequently, the effects due to local surface topography are obtained by a series of measurements of the altitudes, relative to the foot of the instrument, of a regular network of points on the surface, and the use of formulae devised by Eötvös and Schweydar. These formulae are only valid if the variations in the topography in the neighbourhood of the station are not extreme. Where buildings, large cavities or mounds and similar features are present in the vicinity, recourse must be had in addition to other formulae devised by various workers in this field.

The values of U_{xy}, etc. relating to the subterranean anomalies are finally obtained, i.e. the so-called "subterranean values".

Interpretation. The interpretation of the last series of values in terms of the location, magnitude, etc. of geological structures and mineral deposits is a matter partly of geological skill, partly of experience and local knowledge of the probable subterranean features, and partly of the application of rules based on the consideration of the effects, calculated mathematically and expressed both by means of formulae and by graphs, due to various types of "density anomaly" corresponding approximately to actual geological anomalies. The types assumed are generally of regular form whereas the existing anomalies may be very irregular, but we can proceed by successive stages of approximation.

The first stage is to draw up the gravity map of the area, employing the "subterranean values" as found above, but combining them into what are known as "gradients" and "curvature vectors".

We have

$$U_{xz}=\frac{d^2U}{dx\,dz}=\frac{d}{dx}\left(\frac{dU}{dz}\right)=\frac{dg}{dx},$$

which is the rate of change of g per unit distance along Ox, or the "gradient" of g in the direction Ox.

Similarly $U_{yz}=dg/dy=$ gradient of g in the direction Oy, and both are components of the resultant gradient G, which is directed at an angle θ with Ox where $\tan\theta=U_{yz}/U_{zz}$.

Similarly the quantities U_{xy} and U_{Δ} are combined to form a type of vector specified by a magnitude R along a direction making an angle λ with Ox, where $\tan 2\lambda=-2U_{xy}/U_{\Delta}$, which has two solutions λ_1, λ_2 differing by $\pi/2$. The directions λ_1, λ_2 then give the directions of the two principal axes of curvature of the level surface, or potential surface at O.

The gradients G and curvature "vectors" R at each observing station in a

gravity survey are plotted as shown in figure 4, that direction λ_2 being selected for λ which makes sec $2\lambda_2$ of the opposite sign to U_Δ. A pictorial representation of the gravity variations over the region is thus obtained, corresponding, e.g., to a hachured map in a topographical survey. By a simple process, from the gradients, a series of closed lines, known as isogams, can be drawn on the plan. These lines resemble contour lines on an ordinary map, being everywhere at right angles to the gradients, when these latter are corrected by a process of least square adjustment.

The isogam chart is a remarkably effective visual guide to the interpretation of the subterranean anomalies which give rise to the gravity variations. Where the lines are parallel, with the isogams uniformly spaced apart, we can assume that there is a uniform monoclinal type of structure below the surface. Dome-like structures give rise to isogams resembling the contour plan of a surface dome, and so on.

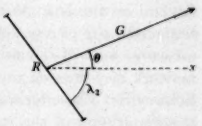


Figure 4.

Other aids to interpretation are the variations of the magnitude of the gravity gradient along a section line of the area. A species of profile chart is constructed for selected sections, and an experienced geophysicist can, from a series of such profiles, make a shrewd judgment as to the subterranean anomalies. The gradients, for example, tend to be very large above steep slopes in the surface separating two media of different density, such as fault planes.

The reduction of torsion balance observations for topographical and other effects, and the computation of the "Eötvös" magnitudes due to assumed solutions of a particular problem of interpretation, have presented problems of graphical and mechanical integration of functions, involving as variables the distances, azimuths and elevations of a set of mass-points, which have led to the development of many ingenious instruments. These studies and the apparatus should have applications in other branches of physics. The graphical methods of Jung, Haalck, Heiland, Below and Lancaster-Jones, and the integraphs of Gamburzeff and Askania Werke, may be mentioned.

Measurement of g itself. Measurements of the intensity of gravity, as distinct from the derivatives of this intensity or its direction, are to-day made almost exclusively by means of the pendulum. On account of the numerous disturbing effects which must be considered in determining the period of oscillation of a pendulum, this measurement must be classified among the most precise ones known. It involves the determination of a period of a half-second with an accuracy of 10^{-7} sec.

Absolute measurements of g are made with the "reversible" pendulum developed by Kater and others. Relative measurements of the difference in g at two places on the earth's surface are made with pendulums of "invariable" length. The intensities of gravity at the two places are thereby inversely proportional to the squares of the periods of oscillation. The pendulum commonly used is the Sterneck half-second one, constructed of invar to afford constancy of length under

varying temperature conditions, and is usually swung in company with one or more others—oscillating in different phases or planes—within an evacuated chamber. For measuring the period of oscillation coincidence apparatus is employed in conjunction with a clock, so that the oscillations of the pendulum and the clock can be compared. The clock is controlled by means of wireless signals. The normal period of observation at a single station is 24 hours.

Observations of the above type have only a modest bearing on the problem of gravitational prospecting, in that they permit a check on the major features of the gravity isogam plan when large concealed structures, such as salt-domes for example, are being investigated. A gravity interval of a few milligals, i.e. units of the order 0·001 c.g.s., is then possible between the regions of minimum and those of maximum intensity of gravity in the area.

Partially with the objective of utilizing direct measurements of the intensity of gravity in applied geophysical problems of this order, attempts have been made in recent years to evolve more portable and rapidly operated apparatus than the standard half-second pendulum outfit employed in geodetic work.

In connexion with such attempts, the early investigations of Threlfall and Pollock in the construction of a quartz-thread static gravimeter are interesting. This instrument has a stretched horizontal torsion wire on which is mounted a light lever having its centre of gravity just off the axis of the thread. The moment of the lever about the wire tends to tilt the lever, but the latter is preserved in a horizontal position by twisting the torsion wire through a measured angle. The instrument, evolved between 1889 and 1899, was never of sufficient sensitivity to make it adaptable to geophysical prospecting.

Among recently developed devices may be mentioned the "static gravity-meter" of Haalck, the Holweck-Lejay "elastic pendulum", and the Thyssen "gravimeter".

The Haalck gravity-meter, devised in 1931, is based on the principle of the differential barometer, and the change in pressure caused by a variation in the force of gravity. According to Haalck, an accuracy of 3 milligals has recently been attained with this instrument, which is easily transported, quick in action, and does not need to be accurately levelled upon a stable foundation.

The Holweck-Lejay pendulum, evolved in 1930, employs a quartz mass attached to a vertical invar strip. The mass is thus controlled by the rigidity of the fibre, and the change in the intensity of gravity causes a variation in the period of oscillation, which is measured. An observation can be taken in about 15 min., with an accuracy of about 1 milligal. The drawback to this arrangement is that the instrument needs a very stable foundation, as it is practically a horizontal seismometer.

The Thyssen gravimeter. Instrumental details of this device, invented in 1934, are not available, but it is claimed that the instrument is simple, light (about 33 lb.), quick acting (observations only take a few minutes) and sensitive to 0·5 milligal. It needs no special protection from temperature, or special care in mounting.

§ 3. MAGNETIC METHOD

The fact has been known for centuries that certain ore-deposits, e.g. of magnetite, distort the normal terrestrial magnetic field on the surface in their vicinity, and influence accordingly the deviation of a compass needle or the inclination of a dip needle. The compass and dip needle are, however, so insensitive that they can only detect the existence of very shallow deposits of highly magnetic ore.

Modern geophysical prospecting is concerned with the detection and location of relatively deep structures or deposits of materials which may be only slightly magnetic. Much useful work can be done in locating deposits of material which is moderately magnetic by means of the somewhat crude instruments of the magnetometer type, such as the Thalen-Tiberg and Thomson-Thalen patterns which were successfully employed in Sweden in the last decades of the nineteenth century. For more precise measurements, the Schmidt horizontal and vertical variometers, devised in 1921, have come into general use. These instruments are sensitive to a few gamma, the unit equal to 10^{-5} of the normal c.g.s. unit of one gauss.

The terrestrial magnetic field. The earth's magnetic field is specified by means of a simple vector, its intensity, which has the customary characteristics of a force in three dimensions, and needs three co-ordinates to define it. The co-ordinates may be polar, i.e. the magnitude and angular directions relative to the vertical and geographical north, or cartesian, i.e. three components, one vertical and two horizontal. Any one of these parameters, or a combination, may be measured by means of suitable instruments, but the most useful one is the vertical component Z of the intensity.

The Schmidt vertical variometer. This instrument consists essentially of a magnet pivoted on a knife-edge so as to be capable of movement only about a horizontal axis. The magnet is weighted to compensate it for the normal vertical intensity and rests nearly horizontal; its equilibrium position moves through a small angle under the influence of slight increases or decreases in the vertical intensity compared with the "normal". The control resisting the tendency of the magnetic intensity to rotate the magnet is partly due to the weight of the magnet itself, since its centre of gravity is below the axis of support and can be adjusted nearer to or farther from the latter by means of horizontal screws; but auxiliary deflecting magnets can be attached as a further control to extend the range of operation and for calibrating purposes.

The magnet actually consists of two similar parallel lamellar bar magnets of tungsten or cobalt steel attached to a quartz knife-edge. When in use the whole instrument is so oriented about its vertical axis of rotation that the magnets can only move in the vertical plane perpendicular to the magnetic meridian. In this position they are acted upon by the vertical component only of the terrestrial field.

A mirror is mounted upon the connecting piece which joins the two magnets, and a vertical telescope mounted above the casing containing the magnets carries a fixed transparent scale. An image of this scale is formed by light, introduced from the side and reflected downwards on to the oscillating mirror, thence upwards to

the eyepiece of the telescope. The direct image of the fixed scale and the reflected image are viewed simultaneously in the focal plane of the eyepiece. The relative displacement of the two images measures the tilt of the magnet, and is indeed directly proportional to the difference between the vertical intensity and the standard "normal" intensity for which the instrument is set. Actually it is the variation of this "incremental" vertical intensity ΔZ which is used as a measure for detecting and locating the buried deposit.

In the early designs of the Schmidt variometers, much trouble was experienced from the effects of temperature, since both the gravity control of the instrument and its own magnetic moment depend on lengths which vary with temperature. In recent years many ingenious devices and modifications of the magnet system have been evolved to compensate for these temperature disturbances.

The Schmidt Horizontal Variometer resembles the vertical instrument in its major details, but has not been much used in practice.

In employing a magnetometer for geophysical work, due account must be taken of variations in the earth's field due to causes which have no relation with the structure or deposit sought. The ordinary diurnal variation of the field, local magnetic storms and other incidental disturbances must be recorded on a separate instrument set up at a stationary base station, unless records can be obtained from an observatory in the vicinity.

Although the Schmidt vertical variometer, or some modification of it such as Haalck's Universal Magnetometer—a combined vertical and horizontal instrument —is the main instrument used in magnetic prospecting, some observers still employ older types such as the dip needle, the Thalen-Tiberg magnetometer and the torsion horizontal magnetometer used by Walker, or modifications of these types. Brief descriptions of these, and of new rivals to the Schmidt variometers, are therefore given below.

Thalen-Tiberg magnetometer. This instrument is still employed for measuring local variations in both the horizontal and the vertical intensity components of the terrestrial magnetic field, when the variations are relatively large, of the order of 200γ to 2000γ. It usually consists of a Tiberg inclinator—comprising a compass which can be set in a vertical plane and oriented so as to be perpendicular to the meridian and which has an auxiliary arm on which can be mounted, at a registered distance, an auxiliary deflecting magnet—provided with an auxiliary Dahlblom arm, slotted to receive a deflecting magnet of which the distance from the compass needle is readily adjustable through a measured extent. By this arrangement the horizontal intensity can be measured according to the well-known tangent or sine methods, or by Dahlblom's sine method. The vertical intensity can be measured by determining the angle of dip in a plane perpendicular to the meridian, and using the predetermined horizontal intensity.

Hotchkiss superdip magnetometer. The dip circle, in which a movable weight, sliding upon one half of the ordinary dipping needle, balances the effect of the earth's field in tilting the needle, is a relatively insensitive instrument, and has a variable sensitivity depending upon the inclination of the field. It has, however,

been a favoured instrument for crude magnetic prospecting in America. The Hotchkiss Superdip is a modern improvement on the ordinary dip circle. By attaching the counterpoise not to the needle itself, but to a supplementary arm pivoted on the same axis as the needle and at an adjustable angle in the plane of the needle, the sensitivity of the instrument can be increased at will by arranging that the supplementary arm shall be more nearly horizontal when the needle is inclined perpendicular to the earth's field, the plane of the needle and arm being in the meridian. The opposing couples due to magnetism on the needle and gravity on the counterpoise are both at a maximum simultaneously when the arm is horizontal and the needle at right angles to the intensity. In practice the arm is arranged to be nearly horizontal when the needle is perpendicular to the intensity, giving an actual sensitivity of the order of about 10γ per degree of inclination.

It is claimed that the superdip provides a rapid, easily operated sensitive instrument, capable of use in any place over any type of magnetic anomaly on account of its readily adjustable sensitivity.

Torsion magnetometers. On general principles, the employment of torsional elasticity instead of magnetic or gravitational control, as a means of counteracting and measuring the magnetic torque in a magnetometer, has many advantages. A torsion wire suspension, as is shown by the Eötvös balance and many magnetic observatory instruments, can provide very high sensitivity, and is largely free from some of the mechanical disturbing effects associated with knife-edges. The forces encountered in magnetic work are, however, much greater than the residual forces operating upon the gravity balance, and there is not the same need for such extreme sensitivity. Nevertheless Walker in 1914 devised a simple portable horizontal magnetometer in which the torsional rigidity of the suspension wire provided the restoring torque operating against the torque due to the earth's horizontal component H acting on the suspended magnet. Walker was satisfied that an accuracy of 5γ in the measurement of H was easily possible. Fordham designed an improved form of Walker's instrument, which was produced by Oertling in 1928. Koenigsberger also devised a vertical variometer having a torsion suspension in 1924.

Electrical-control magnetometer. The magnetic field produced at the centre of a coil of wire in which an electric current is flowing provides another sensitive counter force with which the torque due to the earth's field acting on a magnet can be measured. In the portable horizontal magnetometer designed a few years ago by Sir F. E. Smith and manufactured by the Cambridge Instrument Company, this method of control is employed to produce a very sensitive instrument capable of measuring H to an accuracy of 1γ. The instrument is, to all intents and purposes, a sine galvanometer and its success is primarily due to the accuracy with which an electric current can be measured by means of a potentiometer. A complete determination of H can be made in about 10 min., and the value of H measurable may be anywhere between 0·16 and 0·19 G. with the potentiometer employed. To ensure portability the magnet system is floated on its point in a petrol-filled chamber, the liquid also providing damping. A pair of Helmholtz coils is used to provide the necessary uniformity of magnetic field.

Earth inductor. The opposite principle, of rotating a coil of wire in the earth's field and thereby inducing in the wire a fluctuating current, is employed in the earth inductor to measure the intensity of the field. Magnetic observatories are usually provided with some form of earth inductor, and geophysical measurements have been made in America with the aid of such an instrument.

Rieber vibrator magnetometer. In 1928 Rieber designed an experimental magnetometer on the lines of a string galvanometer but used inversely; the magnetization of a very permeable bar, cut in the centre, was used to cause a thin wire, stretched between the "poles" of the bar and traversed by an alternating current, to vibrate. The wire ceases to vibrate when the bar is rotated so as to be perpendicular to the meridian, and the apparatus can register the direction of this perpendicular to within 0·001° of arc. It was suggested that the instrument could be used to measure intensities by destroying the earth's field by means of a counter field produced electrically and measured on a potentiometer.

Magnetic gradiometers. The measurement of the spatial variations of magnetic intensity at a series of points on the earth's surface is rendered difficult by the fluctuations of the intensity at any one point over a given period of time necessary to take the set of observations in the region. This difficulty is usually resolved by recourse to a second, stationary instrument on which the temporal variations are measured separately and corrections made accordingly. An alternative solution is to measure the spatial gradients of the intensity components rather than the actual magnitudes, this procedure having the advantage that the gradients are practically constant in time, over the small region covered by the survey, and are more definitely due to local structural features than to temporal disturbances. Eötvös designed and used instruments of the torsion type for measuring spatial gradients. Rankine has recently devised a torsion gradiometer, somewhat on the lines of the Eötvös balance, for this purpose, whilst Roman and Sermon have evolved a magnetic gradiometer based on the differential action of two earth-inductor coils. Both instruments are as yet rather in the experimental stage. De Beck has recently advocated the employment of a pair of magnetometers, similar in construction and moved over the field of survey at a fixed interval apart, in order to eliminate automatically the temporal variations. This arrangement obviously results in effect in a measurement of a spatial gradient of the intensity which could more simply be made with a suitable magnetic gradiometer.

Theoretical interpretation of geo-magnetic measurements. In magnetic work, as in all problems of the theoretical interpretation of geophysical measurements, the first step is to study the theoretical fields due to hypothetical magnetized bodies, generally assumed to have simple regular form and to be magnetized uniformly, usually with the poles concentrated at specific points or distributed evenly in specific surfaces. In this way, the characteristic variation of the horizontal and vertical components of the intensity—or of any other magnetic parameter—can be studied for certain typical structures. The field records are then interpreted in the light of their magnetic characteristics, as revealed by the records, and of the local geological conditions.

In the early years of magnetic prospecting in Sweden, Thalen, Tiberg and Dahlblom based their interpretations on the assumption that the ore-body could be regarded as equivalent to an ideal magnet, of which the poles were points separated in distance by from five-sixths to nine-tenths of the total length of the ore-body. The field of such a magnet is well known, and the variation of the intensity components in any plane outside the magnet can be reduced to simple formulae. Carlheim-Gyllenskiold, in 1910, considered the ore to be equivalent to an elliptical cylinder, for which case he calculated the magnetic field and arrived at methods for depth determination which he applied to the famous Kiruna ore-body, in north Sweden. Sundberg in 1924 used ore models of similar shape and assumed linear poles at the two ends of the models.

The assumption most simple for calculation is that of a sphere uniformly magnetized. The magnetic potential in this case at an external point is

$$V = \tfrac{4}{3}\pi R^3 J \frac{\cos\epsilon}{r^2},$$

where R is the radius of the sphere, r the distance of the external point from the centre of the sphere, ϵ the angle made by r with the direction of magnetization, and J the intensity of magnetization.

On the assumption that the sphere is magnetized by the induction due to the earth's field, the horizontal increase of intensity dH, the declination δ and the vertical intensity increase dZ due to the sphere at a point P are given by the formulae

$$dH = -C\frac{H_0}{r^5}\left(t^2 - 2x^2 + 3xt\frac{Z_0}{H_0}\right),$$

$$dZ = -C\frac{Z_0}{r^5}\left(x^2 - 2t^2 + 3xt\frac{H_0}{Z_0}\right),$$

$$\tan\delta = C\frac{3y}{r^5}\left(x - t\frac{Z_0}{H_0}\right),$$

where

$$C = \frac{(\kappa - \kappa_0)\tfrac{4}{3}\pi R^3}{1 + \tfrac{4}{3}\pi(\kappa - \kappa_0)},$$

t is the depth of the sphere below the ground, and x, y are the co-ordinates of P when the origin is vertically above the centre of the sphere, x being in the direction of magnetic north, y of magnetic east. $\kappa - \kappa_0$ is the difference between the susceptibility of the ore-body and the surrounding earth.

An important conclusion from these formulae is that two embedded spheres, having centres at the same depth but possessing different radii inversely proportional to their excess permeabilities, will produce the same external magnetic fields at all points of the surface. Hence an ambiguity of interpretation can arise from measurements made on this surface.

Eötvös derived certain important relationships between the gravitational potential arising from a buried mass and the magnetic potential due to it on the assumption that it was magnetized uniformly by the earth's field. The relationships are derived from Poisson's displacement theorem of the potential. If V denotes

the gravitational potential, at a point P, due to an embedded mass, the gradient of V in a direction i is obtained by imagining the mass translated through a small distance di and computing the resulting change in V. This is equivalent to computing the potentials due to the two shells, formed between the surface of the mass in its old and new positions—all the elements of mass common to the two positions producing contributions to the potential which are ineffective—and subtracting these shell potentials. But these two shells, on the assumption that their elements of mass are replaced in proportion by elements of magnetism represent precisely the magnetic state of the original mass if magnetized uniformly in the direction i. Hence, if W is the magnetic potential of this mass due to the magnetization parallel to i, we have

$$W = \frac{I}{G\sigma} \cdot \frac{dV}{di},$$

I being the intensity of magnetization, G the gravitation constant, and σ the density of the mass relative to its surroundings.

The component intensity in the direction x of the field resulting from the potential W is given by

$$X = \frac{dW}{dx} = \frac{I}{G\sigma} \cdot \frac{d^2V}{dx\,di}.$$

Hence the calculation of the components of magnetic force due to a uniformly magnetized deposit can be derived from the second differentials of the gravity potential, i.e. from those magnitudes which are measured and calculated in gravitational prospecting.

Rigidly, it has been shown that the above reasoning is only valid for a magnetic field arising from induction by the earth's field when the body in which the magnetization is induced is bounded by a surface of the second degree. In practice, however, the formulae derived in gravitational prospecting for calculating the effects due to rectangular plates, wedges, prisms and other "two-dimensional" bodies have been applied to magnetic interpretation on the assumption that the above formula is sufficiently valid.

Owing to doubt as to the precise phenomena underlying the origin of magnetism in buried structures, and the validity of some of the formulae employed—many of them empirical—there is not the same precision in interpreting the records of a magnetic survey as exists in the gravitational case, but sufficiently accurate predictions have resulted in many practical surveys to render the method justifiable.

It has the advantage over the other geophysical methods of prospecting, except the electrical, that it is relatively cheap. The apparatus used is not elaborate or expensive and a large area can be prospected in a short time.

§ 4. SEISMIC METHOD

The beginnings of seismic prospecting can be traced to the development of earthquake recording, with its resulting knowledge of the nature of the earth's crust and the waves of low frequency transmitted through that crust. In the latter part

of the last century Milne developed his horizontal seismometer and similar instruments were evolved in Germany, Italy and Austria. Important advances were made by Wiechert about 1900, especially in the introduction of a satisfactory form of damping. Vertical seismometers were constructed, and the work of Galitzin in developing his electrical seismometers prior to 1914 is outstanding. During the Great War, methods of detecting artificially created disturbances due to gunfire were developed in the various armies, and in particular the work of Mintrop in Germany and the invention of the Tucker hot-wire recorder in England assisted the evolution of seismic prospecting. The real impetus, however, came after the War from the investigation of structures such as salt domes, favourable to soil accumulation, particularly in America, where Fessenden in 1913 had attempted exploration by supersonic methods. In the decade 1920–30 various types of portable mechanical and electrical seismographs were evolved, almost all designed for the so-called "refraction" method of prospecting for fairly shallow structures. In recent years the application of "reflection" methods to deeper seated structures has led to the evolution of somewhat different instruments. Refraction seismometers have as their main function the recording of the exact time of arrival of the first "onset", i.e. the wave which reaches the instrument first. Reflection seismometers are designed to damp down the record of the first onset, which is usually the unwanted direct wave, in order that the arrival of the later reflected wave may be appreciated. In each case the requirements for the field seismographs used in seismic prospecting are very different from those suited to observatory instruments used to record natural earthquakes. The sensitivity of the geophysical patterns needs to be enormous in comparison with the stationary types.

Theory. Elastic, rigid substances, such as the varied materials which compose the earth's crust, readily transmit both longitudinal and transverse vibrations. Seismic prospecting is mainly based on the capacity of the earth material to transmit the longitudinal, low-frequency (infra-audible) vibrations, set up by explosions, with velocities depending upon the compressibility, rigidity and density of the substance. The time of transmission from the origin to a measuring site depends upon the media traversed by the wave. From records of the times of arrival of a wave at a series of observing points, and from an assumed theory of propagation which is supported by practical surveys, deductions as to the nature and position of the intervening media can be made.

It is assumed that in addition to transverse waves and air-borne waves two distinct longitudinal waves are normally in question. The first longitudinal wave is assumed to travel just below the surface of the ground, i.e. practically direct from "shot" point to observing point, with a speed characteristic of the surface medium, or media, if more than one is traversed by this path. The second wave is assumed to travel downwards from the shot point until it meets a transition surface separating the surface medium from an underlying one; if this latter is such that it transmits the wave at higher speed than the upper medium, a refracted wave is created at the point of entry of the initial wave and travels along the separating surface, continually sending out secondary waves which reach the ground surface by paths

similar to that originally proceeding from the shot point. The angles of incidence and emergence of these initial and final waves are each usually assumed to be the critical angle for the two media. In addition a secondary wave is reflected from the interface.

Strict mathematical theory of propagation of waves in non-homogeneous media does not support the above working hypothesis, but denies the possibility of the presumed direct and refracted waves along the paths assumed. However, practical prospecting has demonstrated that the assumption provides working laws which suffice to enable the seismic records to be interpreted correctly, so far as tests can determine.

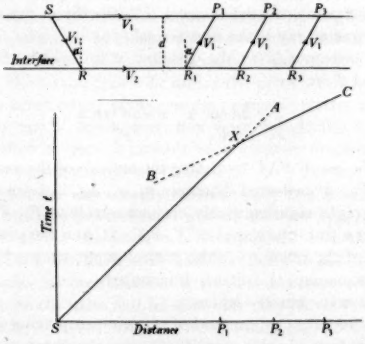


Figure 5.

Working theory of seismic prospecting. Refraction method. This method is specially adapted to the case when a surface medium in which low-frequency seismic waves are transmitted with relatively low speed overlies a medium transmitting such waves with relatively high speed. Such cases are frequent, e.g. as when alluvium rests upon rock. From the shot point S (figure 5) a series of "rays" traverse the upper medium with speed V_1. A direct ray proceeds immediately below the surface, parallel to it, and arrives at the observing points P_1, P_2, etc. at time x_1/V_1, x_2/V_1, x_3/V_1, etc. A second, refracted ray $SRR_1\ldots$ strikes the interface at the critical angle α, and is refracted along the interface in the second medium with velocity V_2, giving rise at R_1, R_2, R_3, etc. to rays which leave the interface, traverse medium *1* along parallel paths, likewise making the critical angle α with the normal to the interface, and arrive at the points P_1, P_2, P_3 in times

$$\frac{2d\sec\alpha}{V_1}\cdot\frac{x_1-2d\tan\alpha}{V_2},\quad \frac{2d\sec\alpha}{V_1}\cdot\frac{x_2-2d\tan\alpha}{V_2},\quad \text{etc.},$$

i.e. denoting these times by t_0+t_1, t_0+t_2, t_0+t_3, etc., where

$$t_0=\frac{2d\sec\alpha}{V_1},$$

we get

$$\frac{t_2-t_1}{x_2-x_1}=\frac{t_3-t_2}{x_3-x_2}=\ldots=\frac{1}{V_2}.$$

If we plot times of arrival of this refracted ray as ordinates against distances from the shot point as abscissae, we get the straight line BXC, inclined at angle $\cot^{-1}V_2$ to the axis SP. Plotting the time-distance graph for the direct ray, we get the line SXA, inclined at angle $\cot^{-1}V_1$ to SP. The broken line SXC thus gives, for any distance (abscissa), the time (ordinate) when the earliest ray to arrive at any point P reaches that point. In the interval SX, the direct ray arrives first, in the interval XC the refracted ray is the first arrival.

The point X corresponds to the abscissa x for which the two rays arrive simultaneously, and therefore

$$\frac{x}{V_1}=\frac{2d\sec\alpha}{V_1}+\frac{x-2d\tan\alpha}{V_2} \qquad \ldots\ldots(1).$$

We can draw the graph SXC by timing the arrivals of the earliest ray at a series of points P_1, P_2, P_3, at measured distances x_1, x_2, x_3, ... from the shot point, by setting up seismographs simultaneously or successively at P_1, P_2, etc. This graph gives us V_1, V_2 and x from the slopes of X, SX, XC and the position of X. Also α is known in virtue of the relation for the critical angle, $\sin\alpha=V_1/V_2$. We can thus determine d from equation (1), suitably transposed.

The equation is $x(1-\sin\alpha)=2d\cos\alpha$.

The theory of the reflection method of seismic prospecting is simple, the echo principle being at its base. In this case, if x denotes the distance of the seismometer from the shot point, we have

$$2d=V_1t\cos\alpha; \quad x=2d\tan\alpha,$$

and d is known as soon as V_1 and t are determined.

As a rule, very small values of α are used so that the discrepancy between the arrival of the direct and reflected waves may be as large as possible.

The theory for determining the depth and inclination of an interface separating a low speed surface medium from a higher speed formation covered by the former, when the interface is inclined to the horizon, and for more complex cases such as a series of media occurring in vertical succession, may be worked out in a similar way. The unambiguous determination of such cases usually involves "shooting" in various directions at different times. Methods of "profile" shooting have been elaborated to cope with problems involving structures such as buried anticlines, salt domes, etc.

Prospecting seismographs. The seismographs normally employed in stationary observatories for recording natural earthquakes are neither sufficiently sensitive nor portable for geophysical exploration. Although many scientific seismologists

had suggested the use of seismic observations to determine geological structure, it was not until about 1923 that practical seismic exploration was initiated, largely in America for detecting salt domes. The Mintrop seismograph, operated by the Seismos Company, may be regarded as the practical pioneer instrument.

Seismographs may be differentiated according to several modes of operating. The usual method in earthquake work is to differentiate between horizontal and vertical instruments, the former registering only the horizontal and the latter only the vertical component of the earth tremor. In geophysical prospecting, seismographs are more usually distinguished according to the dynamical function of the tremor to which they react. Thus we have displacement, velocity and acceleration types. The first are usually mechanical, vertical component seismometers; the second are electromagnetic induction seismometers, the original displacement of the ground being transformed by means of electromagnetic induction into a variation of current which is proportional to the velocity of the motion, not to the actual displacement. The third type generally utilize the principle that a variable force is proportional to the acceleration of the motion related to it; the displacement of the ground is made to vary a pressure exerted upon an electrical device, such as a microphone, and thus to vary a current in a manner proportional only to the acceleration of the original ground movement.

The last two types are also usually sensitive only to the vertical displacement of the ground. Mintrop's seismograph was of the first type, and recorded the vertical component of the displacement. It consisted of a mass, about 20 lb., attached by means of a flat horizontal lamellar spring to the massive base of the apparatus and capable of oscillation in a vertical direction. A light metal cone rested with its base on the mass and its apex vertically above the mass. To the end of the cone was attached a horizontal spindle, which rubbed by friction against a pivoted vertical spindle carrying a mirror. The arrangement of cone and spindles served to transform the vertical displacement of the mass relative to the main base frame into a rotation of the mirror about the vertical spindle, and also to magnify the motion. Further magnification, up to several thousands in all, was obtained by an optical lever, in which light impinged on the mirror and was reflected therefrom on to a distant sensitive paper, which gave a magnified photographic record of the disturbance.

Schweydar in 1926 designed a modified Mintrop seismograph in which the connexion from the end of the cone to the spindle of the mirror was made by a thin wire, kept in tension by a spring and wound once round the spindle. Schweydar's seismograph had two units, for recording horizontal and vertical displacements respectively, and its magnification was about 20,000.

The movement in these seismographs was damped by liquid or electromagnetically, a thin metal plate attached to the top of the cone vibrating in the field of a permanent magnet.

J. H. Jones, in conjunction with Messrs Griffin and Tatlock of London, has designed a modification of Mintrop's seismograph to ensure enhanced, and adjustable, sensitivity, the magnification attaining 50,000 at the limit. Over the upper

end of the cone is stretched a thin strip carrying a small soft iron plate in the field of a pair of permanent U-shaped magnets. The position of these magnets is adjustable so that the magnetic lines in the vicinity of the plate can be arranged to ensure that, when the plate vibrates with the earth tremor, it tends to rotate under the influence of the field. The rotation is proportional to the displacement. A mirror is attached to the plate.

The Cambridge Instrument Company has also made a mechanical seismograph of design similar to Mintrop's, except that the cone is inclined almost horizontally, and operates the reflecting mirror through a magnifying intermediary arrangement consisting of a pivoted needle impinging upon the periphery of a disc attached to the mirror spindle. In this instrument, the simple leaf spring of Mintrop is replaced by a pair of crossed leaf springs, and the movement is damped by means of springs on the Galitzin principle.

By substituting a small coil, operating in the field of a permanent magnet, on the Galitzin principle, for the mechanical amplifier of these displacement seismometers, it is easy to transform them into velocity seismometers of the induction type. In the simple induction type, even the conical amplifying element is omitted. The Cambridge Instrument Company's pattern retains this feature and uses the coil only as the terminal element of its mechanical amplifying linkage. Other patterns have been designed by Karcher, the Askania Company and the Imperial Geophysical Survey. The Askania model is a modified Schweydar mechanical seismometer. In the seismometer of the Imperial Geophysical Survey, an electromagnet acting as the inert mass is suspended from a thin steel diaphragm. Attached to the base, and passing between the poles of the electromagnet, is a flat stationary coil.

In reluctance types of seismometer, a soft iron armature is attached to the inert mass and moves in the field of a permanent magnet to which coils are attached; by the change of flux, induction currents are produced in the coils.

Seismometers of the capacitative type have been designed, the mass being mounted in close proximity to a stationary plate, to form a condenser of capacity varying with the seismic oscillations. The condenser can be used to change the tuning of an oscillatory circuit or to change the grid voltage of the first stage of a valve amplifier, as in a condenser microphone.

The principle of the microphone lies at the basis of many designs of prospecting seismometers. The carbon microphone has not proved very satisfactory. The Imperial Geophysical Survey also designed a hot-wire seismometer using the Tucker microphone and therefore acting as a velocity seismometer. Ambronn produced a seismograph of the contact-accelerometer type and several piezo-electric seismometers have been designed to provide accelerometers, notably by Galitzin, Wood, Kato and Rieber.

In all cases where the seismometer is used to provide a variable current, valve amplification and oscillograph recording provide the most effective auxiliaries for field use. The amplification can be controlled in order to obviate the disturbing effects due to ground unrest and wind, and the oscillograph lends itself to the employment of a number of seismometers, one at each of several stations along a

traverse line, recorded on a single film. The Einthoven string galvanometer, as used by the Cambridge Instrument Company, is an efficient recording mechanism.

It is also possible to bury the seismometers some feet below ground level to avoid disturbance from ground and air currents, and in many designs the instrument can be used below water with enhanced effect. Owing to the valve amplification, a certain amount of frequency tuning can be introduced to discriminate between useful and useless vibrations.

§ 5. ELECTRICAL METHODS

Electrical methods of prospecting involve the various electrical and magnetic characteristics of rocks and minerals, and the complete range of frequencies of the electric currents or electromagnetic oscillations applied to the ground, including as a particular case direct current, although the d.-c. method has become almost obsolete on account of the difficulties due to polarization of the electrodes whereby the current is introduced to the ground, and due also to the presence of stray natural earth currents. The principal electrical characteristic of rocks and minerals concerned is the conductivity, or its inverse, the resistivity. Natural rocks *in situ* vary enormously in this respect, usually on account of their porosity, their situation in respect of the natural water table of the district, and the chlorine content of the permeating water. For example, dry sand is highly resistant, but sand saturated with salt water is a very good conductor of electricity. Mineral ores, even if metallic, may not be good conductors but the ore-minerals which possess a metallic lustre, such as the sulphides of copper, iron and lead, are with few exceptions good conductors. The conductivity of minerals is often seriously impaired by disseminations of gangue, whilst rocks are frequently anisotropic in their conductivity characteristics.

The three principal methods are known respectively as spontaneous-polarization, surface-potential and electromagnetic-field methods. In the first named, natural currents arising in virtue of electrochemical action in certain deposits—such as the oxidation processes occurring when certain sulphide ores lie near the surface—produce at the surface a field of electric potential which can be tapped by means of electrodes connected to a galvanometer, milliammeter or potentiometer circuit. Lines of equipotential and profiles of potential variation can be set out on the ground. The region directly above the deposit is a centre of negative potential. Special non-polarizable electrodes must be used.

Surface-potential methods involve the introduction into the ground of an artificial current, either direct or alternating, by means of supply electrodes. Where the ground is uniform, the lines of current flow and equipotential lines present a regular character depending upon the arrangement and number of electrodes used. If two "point" electrodes are used, the current lines correspond to the lines of force due to a single point source and a single point sink in a homogeneous field, i.e. the well-known arrangement of magnetic lines of force for a simple bar magnet. The presence of a relatively good conducting mass below a part of the region traversed by the current distorts the field in that region, causing the

current lines to bunch together, and the equipotential lines to diverge from one another. The lines are mapped out on the ground by means of a circuit similar to that used in the spontaneous-polarization method if direct current is supplied to the ground. If the supply is alternating, at audible frequencies, a simple search circuit consisting of a pair of iron probe electrodes connected through a telephone headpiece may be used. The current picked up by the phones may be amplified by a single, two-stage or multi-stage amplifier to make the method more sensitive. Some difficulty arises from the phase variation effects in the presence of good conductors, when alternating current is used. Secondary induced currents arise in the vicinity of the conductor, and are out of phase with the primary supply currents. In fact, modern methods make valuable use of these phase effects to detect the good conductors.

In the surface-potential method, potential profiles may be surveyed, but the more usual method is to employ a bridge to provide a potential ratio profile, or a species of ohmmeter to give a resistivity profile.

The electromagnetic-field method involves the measurement, at a series of points just above the surface of the ground, of the magnetic field characteristics produced when alternating currents are supplied to the ground conductively, as in the surface-potential methods, or inductively by means of horizontal or vertical transmitting loops, insulated from the ground. A special case consists in the inductive supply of very high-frequency currents, including radio frequencies of all wave-lengths. These latter are frequently claimed to have special value in detecting very small deposits, and also to be appropriate in arid regions where currents of lower frequency are ineffective, but the practical value of the high-frequency methods still remains somewhat uncertain. Electromagnetic-field methods again introduce the phase complication, and here also the measurement of the out-of-phase component of the secondary field arising from a local, buried conductor is a valuable aid in detecting and locating the latter. An important factor is the frequency of the supply. To some extent the higher frequencies are advantageous in eliciting a high response factor in the conductor sought, provided too much energy is not absorbed by the intervening earth.

Theory of electrical methods. The distribution of currents and the corresponding potential function for the normal earth, consisting as this does of a series of irregularly shaped media of different electrical characteristics, is a problem to which only approximate theoretical treatment, based on the hypothesis that the media are bounded by simple surfaces, is possible. However, the following simple case, for which an exact theoretical solution is obtainable, represents with sufficient accuracy for practical purposes a number of important methods.

Let XX' (figure 6) represent the intersection of a horizontal plane which separates a medium of resistivity ρ from the air, for which the resistivity is taken to be infinite. The medium ρ is assumed to extend to infinity in all directions below the plane XX'. At points A, B respectively in the lower medium are a source I and an equal sink $-I$ of current. The current will flow from A to B through the lower medium. If A_0 and B_0 are the respective images of A and B in the plane XX'

and distances measured to A, B, A_0 and B_0 respectively are denoted by r, s, r_0, s_0, the potential V in the medium ρ is known to be given by the equation

$$V=\frac{I\rho}{4\pi}\left(\frac{1}{r}-\frac{1}{s}+\frac{1}{r_0}-\frac{1}{s_0}\right).$$

This represents a potential satisfying Laplace's equation in the medium ρ, except at points A and B, where it satisfies the conditions of a source I at A and a sink $-I$ at B. It also satisfies the condition that there shall be no flow of current across the plane XX' and that the potential at infinity shall be zero. It is therefore the correct expression for V.

Let the projection of A on plane XX' be L and that of B be M. At L the potential V is obviously positive, since r and r_0 are each less than s and s_0. At M the potential is negative, and very much greater numerically than at L, since at

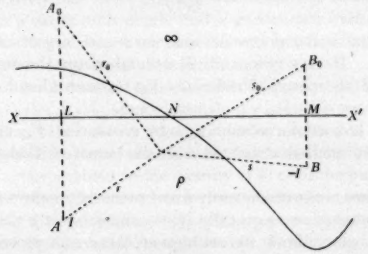


Figure 6.

M, s and s_0 are both very small compared with the values r and r_0 at L. At a point N between L and M on the line XX', V becomes zero, as it is also at infinity. Since $AN=BN$, N is nearer to L than M. The potential at points along the axis XX' is thus represented by the curved line shown, which has a small maximum near L and a larger minimum near M.

This represents quite reasonably well the practical case of a natural earth current arising from an inclined linear ore-body of which the upper end B is oxidized, causing current to flow along the line BA inside the body, and return from A to B through the surrounding medium. V represents the potential of the return current in the medium. If the body is much inclined to the horizontal, the maximum and minimum of the potential do not occur at points of the surface above the ends, but are outside the ends of the body. If the lower end of the ore mass is deep compared with the upper, the potential maximum practically disappears and we are left only with a minimum near the point vertically above the upper extremity. The symmetry or asymmetry of the rise of potential on either side of this minimum gives a measure of the inclination of the axis of the body to the vertical.

In a similar way, the potential field for a polarized thin conducting lamina can be evaluated, the lamina being regarded as consisting of a linear source of current at its lower edge and an equal linear sink at its upper edge. The variation of potential on the surface of the containing medium is similar in character to that for a linear conductor.

The method of images used in these problems is further applicable to the theoretical solution of more complex problems involving successive horizontal layers of media of different electrical characteristics. Maxwell indicated the method of solution in such cases, and Hummel and others have worked out interesting cases for problems connected with geoelectrical prospecting.

When the source A and sink B of the previous problem both lie in the upper surface XX' of the conducting medium, the potential is given by the formula

$$V = \frac{I\rho}{2\pi}\left(\frac{1}{r} - \frac{1}{s}\right),$$

and the field is the well-known one associated with a pair of equal, dissimilar electrical charges. If two points C, D are taken on the line AB, such that $AC = CD = DB = b$, the potential difference V_{23} between C and D is easily seen to be given by the expression $V_{23} = I\rho/2b$, or $\rho = 2\pi b V_{23}/I$.

This formula is used for obtaining ρ by measuring V_{23} and I in the four-electrode resistivity method associated with the names of Gish and Rooney, who employed it in 1925.

Gish and Rooney, and subsequently many practical geophysicists, assumed that, when the medium was not electrically homogeneous, but ρ varied from place to place, the formula was valid on the assumption that ρ now represented the average resistivity of the ground within a region of lateral and vertical distance from the four electrodes equal to the electrode interval b. On this assumption, they varied both the electrode spacing interval b and the position of the set of electrodes in plan, and thereby obtained a series of values for ρ which were taken to correspond to the ground to depths b below the centre of the four electrodes in each position.

Hummel and others have shown that this assumption of "depth of penetration equals electrode spacing" is theoretically not valid, but in practice it affords a rough rule for interpreting such four-electrode resistivity surveys. If the centre of the electrode set is kept constant and b gradually increased, the corresponding resistivity-spacing curve shows more or less abrupt changes in the magnitude of ρ for values of b corresponding to the depths below the surface of interfaces between different conducting layers.

A variant on the four-electrode method consists in the provision of a fifth central electrode E midway between C and D. The potential differences V_{CE} and V_{ED} are then measured separately. This method gives a better indication when the set is moved over a steeply inclined interface such as a fault plane, between two media of different resistivities.

These methods are obviously very useful in obtaining the conductivity coefficients for specimens of ground *in situ*, and afford valuable data in this respect for

subsequent detailed work in which a knowledge of these coefficients is needed. For example, a set of electrodes can be inserted into an outcrop of the natural rock, and the correct resistivity obtained for the area in which this rock subsequently descends below the cover of alluvial deposits, or other rocks.

If the electrode B is displaced to a considerable distance from A, whilst C and D are kept at relatively small distances b, $2b$ from A, we have in this case $2\pi V_{23} = I\rho/2b$ and $\rho = 4\pi b V_{23}/I$. This is the formula for the so-called "single-probe" method of resistivity surveying, which is another variant on the four-electrode method, and has some advantages.

In other variants, the potential difference is measured on the extension of the line BA, on the other side of A from B.

Equipotential-line method. Theory. In the preceding, the potential field due to a pair of current electrodes inserted into the surface of a semi-infinite conductor of uniform conductivity has been obtained. The resistivity method can be used to determine whether ρ varies with depth, and is particularly useful in locating horizontal or nearly horizontal interfaces between widely extending beds of different structure, and also for detecting faults separating structures.

The presence of more compact masses, embedded in the earth at shallow depths, is more readily indicated by a potential line survey in which the equipotential lines are plotted as follows. One of two search probes or electrodes, connected to each other through a potentiometer or similar indicator, is set into the ground at a point P, and the other is moved about in the vicinity of P until the indicator registers no potential difference when the second or forward electrode is inserted at a point Q; the back electrode is then advanced to Q and the process repeated to find a further point R on the same equipotential line as P and Q. In this way the line is gradually mapped out. Similarly neighbouring lines can be plotted. The normal potential field will be that due to the single pair of current electrodes which have remained fixed. Any bunching together or diversion of these equipotential lines from the normal, indicates that in the region of disturbance there lies some mass of conductivity different from the normal. Some idea of the extent and depth of the mass is given by the extent and strength of the dislocation of the equipotential lines, but a more intensive examination of the neighbourhood can be made by resistivity or other measurements.

In a variant of this method, linear instead of point electrodes may be used to introduce the current into the ground. The normal potential field in this case consists of a series of lines parallel to the two electrode lines, which may consist of a series of single electrodes joined together.

In all these methods, measurements should not be taken in the neighbourhood of the electrodes, or of the leads from these to the generator, as in these regions the electrical field is not normal.

The theoretical problem of a conducting sphere embedded in an infinite medium of different conductivity in which a uniform current is flowing is a well-known problem in text-books on electricity, and is capable of precise solution. It corresponds approximately with the geophysical problem of a similar conductor below

the surface when the supply electrodes used are linear, or when, if point electrodes are used, the conductor lies somewhere in the centre of the field, where this is then reasonably uniform if the dimensions and depth of the conductor are small compared with the electrode spacing. For this case the potentials V outside the sphere and V' inside it are given by the formulae

$$V = -\frac{i_0}{\sigma}x - \frac{\sigma-\sigma'}{2\sigma+\sigma'}\cdot\frac{i_0}{\sigma}\cdot\frac{a^3}{r^3}x; \quad V' = -\frac{i_0}{\sigma}x - \frac{\sigma-\sigma'}{2\sigma+\sigma'}\cdot\frac{i_0}{\sigma}x,$$

where i_0 is the current in the undisturbed, homogeneous field of conductivity σ, σ' is the conductivity of the sphere, a the radius of the sphere, r the radial distance to the observation point, and x the abscissa parallel to the direction of i_0. The factor r^3 in the denominator of the second term of the expression for V shows that the effect due to the introduction of the sphere diminishes as the cube of the distance. Thus, as the effect is measured at the surface of the earth, it diminishes very rapidly with the depth of the conducting sphere. The factor $(\sigma-\sigma')/(2\sigma+\sigma')$ also entering into the disturbing effect shows the influence of the ratio of the conductivities σ, σ'. If σ' is very large compared with σ, this factor equals -1, as compared with a value $+\frac{1}{2}$ when σ is very large compared with σ'. On the other hand, if $\sigma' = 10\sigma$, the factor becomes $-\frac{3}{4}$ and is nearly as large as the case when $\sigma'/\sigma = \infty$.

Thus, the surface-potential method does not adequately emphasize the effect of very high conductivity in the deposit sought.

For quantitative work, it is necessary that the relative potential of each line should be obtained by the use of a potentiometer or some simple bridge arrangement. Actually the lines are normally plotted irrespective of potential difference, and the survey is subsequently amplified by taking a few potential profiles along traverses which the line survey have shown to be of particular interest.

Alternating current methods. The preceding considerations only apply rigidly when direct current is supplied to the ground through the electrodes. This has the disadvantage that the search electrodes must be of the rather inconvenient non-polarizable type, of which several patterns exist. The supply of alternating current, on the other hand, introduces a new complication in that the secondary currents induced in conductors traversed are out of phase with the primary current supplied. Consequently the resultant field at the surface may be of the elliptically polarized type obtained by the combination of two alternating currents not in phase with each other. If therefore two search electrodes are inserted into the ground, the current in the flying circuit will practically never be zero, unless secondary currents are not present, but will alternate between a maximum and a minimum which correspond with the major and minor axes of the representative vector ellipse in the horizontal plane.

Thus it will not be possible, at any rate in that part of the surface which is influenced by secondary currents, to plot lines of equal potential, but only bands of minimum potential. The difficulty may be overcome by introducing into the flying circuit a balancing current, 90° out of phase with the primary current and

180° out of phase with the secondary current, so that the latter is rendered inoperative by a suitable variation of the magnitude of the supplementary component. In this way, so-called equipotential lines may be plotted, and a correspondence with the direct current method obtained.

An alternative and preferable method is to measure the out-of-phase secondary component, since this is directly associated with the conducting deposit of which the position and extent are the objects of the survey.

One arrangement for effecting this measurement is the A.C. Ratiometer devised by Broughton Edge in 1929 and used in the Imperial Geophysical Survey. The arrangement, see figure 7, consists of an alternating bridge in which *A*, *B* and *C* are ground electrodes and *AO*, *CO* are the ratio arms in each of which a condenser and resistance are connected in series, whilst a telephone receiver, with amplifier, is in the arm *OB*.

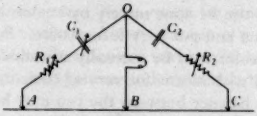


Figure 7.

The capacity in arm *OC* is fixed, but that in arm *OA* and both resistances are variable. They are adjusted until an exact balance, indicated by silence in the telephone receiver, is obtained, when the following relations will hold:

$$V_2/V_1 = R_2 \sin \tan^{-1}(R_1/X_1)/R_1 \sin \tan^{-1}(R_2/X_2)$$

and

$$\theta_2 - \theta_1 = \tan^{-1}(R_2/X_2) - \tan^{-1}(R_1/X_1).$$

This latter equation gives the phase difference between V_2 and V_1.

If a traverse commences in the undisturbed portion of the field, the initial phase angle θ_1 is zero, the ground current being the undisturbed primary current. As the points *A*, *B*, *C* are gradually advanced, at each stage *A* becoming the previous *B*, *B* becoming the previous *C* and *C* becoming a new advance point along the traverse, the phase variation along the traverse is gradually obtained by summation of stage increments. Similarly the potential is obtained by progressive multiplication of the stage ratios V_2/V_1.

In a profile the potential and phase variation curves tend to be complementary, and either can be used to interpret the records in terms of conductors below the surface.

A similar phase compensating meter was developed by Sundberg in 1925.

§ 6. ELECTROMAGNETIC METHODS

Instead of surveying the potential and phase characteristics of the currents at the surface of the ground, the characteristics of the electromagnetic field in the region above the ground may be explored. For such a field to exist, there must be

alternating currents, either introduced into or induced in the ground. The electromagnetic-field survey measures characteristics which depend not merely on the conductivity of the ground, but on its inductive capacity and on the frequency of the alternating current supply. The frequency may be varied at will to suit local conditions, and currents may be set up by induction in the conductors buried in the ground even if the latter is itself non-conducting. Whatever method is used to set up the currents in the buried conductors, the field measuring apparatus is the same. It must be designed to measure the complex elliptically polarized magnetic field which arises from the superposition of the primary and secondary fields due to the original current and any currents induced in local conductors. A single exploring coil in series with a detector can only measure one feature of the field, namely the plane of polarization, since when the coil is in that plane no response will be obtained in the detector.

Although useful work can be done merely by exploring this plane with a single coil, the method is tedious and not very informative. Better results are obtained by using a pair of coils, which may be mutually adjustable as regards their planes, and may also be provided with means for varying their individual inductances and capacities, in order that a balance between the two may be obtained, somewhat on the lines of the a.-c. bridge or inductance meter. Various arrangements have been employed by geophysicists, suited to varying methods of supplying the primary current and focusing upon particular aspects of the resultant electromagnetic field.

One frequently used arrangement is known as the horizontal loop method. In this, the primary current is supplied inductively, at a fairly low frequency of about 500 cycles, by means of a larger insulated loop conductor which forms a circuit of a mile or so around a region under prospect. The loop is supported just above the ground and energized by an a.-c. petrol generator. In the central region within the loop, the primary electromagnetic field is practically vertical. Hence a vertical exploring coil will experience no induction from this field, but only from secondary fields arising from conductors in the neighbourhood. The currents in these conductors will be out of phase with the primary current, the amount of phase difference depending on the electrical characteristics of the conductors and the surrounding ground. The secondary currents will be in the nature of eddy currents, flowing for the most part round the boundaries of the conductors and giving rise to secondary electromagnetic fields which at the surface have both vertical and horizontal components. The vertical secondary component will also not affect the vertical exploring coil, which will only respond to the horizontal component of the secondary field. A second coil, fixed at right angles to the vertical coil, will respond to the resultant of the vertical component of the secondary field and the original primary field. The former will usually be rather small compared with the latter, and as a first approximation we may say that the horizontal coil will be affected by the primary field, whilst the vertical is affected only by the secondary field. We may also assume that this latter is nearly 90° out of phase with the former, and we may introduce a condenser into the circuit of the vertical coil to change the phase of the current induced in it by this amount. The inductance of the

horizontal coil can be varied by putting any desired number of its turns into circuit. If now the vertical coil current is opposed to the horizontal coil current in a detecting circuit, a condition of balance may be obtained by varying the number of turns in the horizontal coil. The number of turns required to be in circuit in order to effect the balance will be a measure of the strength of the horizontal secondary, out-of-phase field as compared with the vertical primary in-phase field. In practice the balancing measurement is made with the plane of the vertical coil first in the north-south plane and next in the east-west plane, and the values obtained represent the north-south and east-west components of the horizontal out-of-phase field. These are combined vectorially to give the resultant out-of-phase field. The magnitude of this field is a useful guide to the proximity of a buried conductor.

The above method is one associated with Bieler and Watson. A variety of methods of exploring the secondary fields exists for electromagnetic as for conductive or surface-potential methods. Among the earlier pioneers may be mentioned Lundberg and Sundberg, who have devised various sets of apparatus and operated electrical methods of both conductive and inductive types, mostly using low-frequency currents.

High-frequency inductive methods. The use of high frequencies, from 10,000 cycles to radio frequencies, has many adherents. Whilst the principles of the high-frequency methods resemble those of the inductive methods, the apparatus is different and certain aspects are more prominent. The generator of the inductive field is now a valve arrangement, with a frequency control. The sending loop is usually quite small and may be vertical or horizontal. The receiving coil or frame is tuned to the transmission frequency, but may also be tuned to neighbouring frequencies, like a wireless receiver. The directional properties of the frame are enhanced by the higher frequencies used, and may be further enhanced by some system of screening, whilst the transmitter can also be made directional, as in beam wireless. The location of buried conductors may be effected through measuring the distortion in the direction of the normal electromagnetic field caused in their neighbourhood by the secondary currents excited in them, or it may be effected by measuring the frequency change in the resultant field in their neighbourhood. Other methods involve the reflection of electromagnetic waves at the surface of conductors, a phenomenon peculiar to higher frequencies. The penetration of the waves of higher frequency is enhanced by high resistivity of the ground, and correspondingly reduced by absorption when the ground conducts well. The high-frequency methods are therefore very useful in arid districts, or in cases where a buried conductor is surrounded by a region of very high insulation.

Electrical coring. A resistivity method of utilizing existing uncased drill holes to explore the strata surrounding them has been developed by C. and M. Schlumberger since 1928 and is known as electrical coring or electrical drilling. The drill hole is usually filled with water, which has salts in solution and makes good contact with the adjacent strata at various depths. The method consists of a four-electrode resistivity apparatus in which one of the supply or potential electrodes is sunk into

the surface of the ground near the mouth of the drill hole, whilst the other three electrodes are lowered by means of a cable into the water in the drill hole. These three are fairly near one another, the second supply electrode being the lowest, and the distances being measurable by means of the cable. It is assumed that the potential surfaces surrounding the lower electrode, due to the supply current through the cable returning *via* the surrounding strata, are approximately spheres, which are tapped at distances r, r' from A by means of the intervening current electrodes. The formula for the resistivity of the stratum between the two current electrodes is then

$$R = 4\pi \frac{\Delta V}{i} \cdot \frac{rr'}{r'-r}.$$

The voltage difference ΔV between these electrodes is measured on a potentiometer at the surface, and the supply current i is measured on an ammeter. The cables carrying the electrodes are then raised or lowered to obtain new values for R at a different depth. Usually the three lower electrodes are kept at invariable distances apart.

REFERENCES

Geophysical Prospecting, 3 vols. (1929, 1932, 1934). American Institute of Mining and Metallurgical Engineers.

Wien, W. and Harms, F. *Handbuch der Experimentalphysik*, **25**, Teil 3, Angewandte Geophysik. Ed. by G. Angenheister (1930).

The Principles and Practice of Geophysical Prospecting. Ed. by A. B. Broughton Edge and R. B. Laby (1931). Imperial Geophysical Experimental Survey.

Shaw, H. *Applied Geophysics* (1931). Board of Education. Science Museum.

Eve, A. S. and Keys, D. A. *Applied Geophysics*, 2nd ed. (1933).

Haalck, H. *Lehrbuch der angewandten Geophysik* (1934).

RADIO EXPLORATION OF UPPER-ATMOSPHERIC IONIZATION

By E. V. APPLETON, F.R.S.

§ 1. INTRODUCTORY

In the absence of measurements made *in situ*, such as are possible for the lower strata of the atmosphere, information concerning the nature of the upper atmosphere is derived from ground observations on (*a*) terrestrial magnetic variations; (*b*) manifestations of luminosity such as are associated with aurorae, meteorites, etc.; and (*c*) radio wave reflection. Although (*a*) provided the first indication that the upper atmosphere constituted a relatively highly conducting shell, which is often termed the ionosphere, the prosecution of exploration by (*c*) has proved, on the whole, the most fruitful. Wireless methods possess the great advantage that an exploration may be made at any time and it is not necessary to wait for natural sequences or irregularities. For this reason the present report is concerned almost wholly with results obtained by radio wave investigations, evidence from the other sources mentioned being included only where supplementary or confirmatory.

Wireless exploration consists in sending waves (usually) vertically upwards and noting the characteristics of the waves reflected by the ionosphere. The quantities immediately measurable in such cases are (1) the group-time of flight of the waves on the up and down journey, (2) the state of polarization, and (3) the intensity, of the returned waves. The phase-time of the waves on the up and down journey is not immediately measurable but, from observations made over a period, the temporal rate of change of this quantity may be determined. For the case of propagation at oblique incidence relative to the reflecting stratum the angle of arrival of the reflected waves at the ground may also be measured. As is shown below, all these quantities may be used in deriving information concerning the electrical state of the upper atmosphere.

§ 2. THE RELEVANT ELECTROMAGNETIC THEORY

The early discussions of the propagation of electric waves through an ionized medium by Eccles[1] and by Larmor[2] took no account of the influence of the earth's magnetic field. Appleton[3], and also Nichols and Schelleng[4], pointed out that, if the effective electrical carriers are of electronic mass, the influence of the earth's field is very important and must be taken into account. The magneto-ionic theory, as the revised treatment is called, has been developed mathematically for the general case of propagation at any angle to the earth's field by Breit[5], Appleton[6], Goldstein[7], Hartree[8] and Baker and Green[9], neglecting absorption, and by Appleton[10] taking into account the absorptive effect of electron collisions with the

gaseous molecules. The general propagation formulae derived by the last-mentioned author have been elucidated graphically by Mary Taylor[11] and by Booker[12], while a detailed general study of the whole theory has been given by the latter.

Only final formulae are cited here. Given the nature of the medium, we require to know the refractive index μ, the absorption coefficient κ and the state of polarization for the two component waves which travel in the doubly-refracting medium. These quantities are expressed in the two fundamental magneto-ionic formulae (1) and (2) for the case of propagation of plane waves (magnetic vectors H_y and H_z) along the axis of x, the imposed magnetic field having components H_L and H_T along x and z (longitudinal and transverse respectively) but not along y (see figure 1):

$$c^2q^2=\left(\mu-\frac{ic\kappa}{p}\right)^2=1-\frac{2p_0^2}{2(p^2-ip\nu)-\frac{p^2p_T^2}{p^2-p_0^2-ip\nu}\mp\sqrt{\left[\frac{p^4p_T^4}{(p^2-p_0^2-ip\nu)^2}+4p^2p_L^2\right]}} \quad \ldots\ldots(1),$$

$$\frac{H_y}{H_z}=\frac{ipp_L(c^2q^2-1)}{p_0^2+(ip\nu-p^2)(1-c^2q^2)} \quad \ldots\ldots(2).$$

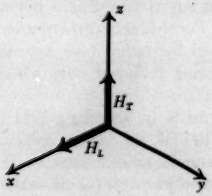


Figure 1.

In these equations

$$\frac{4\pi Ne^2}{m}=p_0^2, \quad \frac{H_Le}{mc}=p_L \text{ and } \frac{H_Te}{mc}=p_T \quad \ldots\ldots(3),$$

where the medium contains N electrons (of charge e and mass m) per c.c., and p and ν are respectively the electric wave angular frequency and the frequency of electron collisions with gas molecules.

Equations (1) and (2) show that there are propagated, in general, two waves of different phase velocity, absorption and polarization. The wave corresponding to the lower sign in (1) is called the ordinary wave since its characteristics are not relatively very much influenced by the imposed magnetic field. (It corresponds, moreover, in the case of purely transverse transmission, to that component for which the electric vector is parallel to the imposed field.) The wave corresponding to the upper sign is termed the extra-ordinary wave since it exhibits the phenomenon of quasi-resonance with the characteristic electron gyratory frequency conditioned by the magnitude of the earth's magnetic field. The ordinary ray is, for the greater part of the wireless spectrum used in these investigations, less absorbed than the extra-ordinary ray. Similarly, the polarization of the ordinary

wave is left-handed and that of the extra-ordinary wave right-handed for a direction of propagation making an acute angle with the positive direction of the magnetic field. Here the polarization is viewed as by an observer looking in the direction in which the waves are travelling.

It is of interest to note from equations (1) and (2) that the medium is isotropic for the special conditions $p^2 = p_0^2$ and $\nu = \frac{p_T^2}{2p_L}$.

Fortunately, certain simple deductions may be made from the rather complicated general formulae. If we neglect friction it is easy to show[6] that the refractive index μ is reduced to zero when $p_0^2 = p^2$ for the ordinary ray, and when $p_0^2 = (p^2 - p\sqrt{p_T^2 + p_L^2})$ for the extra-ordinary ray. Now $\mu = 0$ is the condition determining reflection at vertical incidence. We can therefore specify the value of the electron concentration N at which the waves are reflected in either case[13]. The appropriate relations are, by substitution from (3),

$$N = \frac{mp^2}{4\pi e^2} = 1{\cdot}24 \times 10^{-8} f^2 \quad \text{ordinary ray} \qquad \text{.....(4)}$$

and $$N = \frac{m}{4\pi e^2}(p^2 - pp_H) = 1{\cdot}24 \times 10^{-8}(f^2 - ff_H) \quad \text{extra-ordinary ray} \text{.....(5)},$$

where f is the electric wave frequency and f_H is the characteristic gyro-frequency of electrons under the influence of the earth's total field. For south-east England its value is about $1{\cdot}32 \times 10^6$ cycles per second.

Appleton[14] has employed equations (4) and (5) in the critical penetration frequency method of measuring the maximum ionization content of any particular ionized region. If the electric wave frequency used in the vertical incidence experiments is gradually increased until either the ordinary waves or the extra-ordinary waves just penetrate the region in question, the maximum value of N is found by the insertion of the appropriate critical penetration frequency in (4) or in (5).

Another useful result follows from the fact that the relation between μ and N, for the ordinary wave and for frequencies greater than the gyro-frequency mentioned above, approximates very closely to the corresponding relation when the imposed magnetic field is zero. In such cases we have from (1), putting $p_T = p_L = 0$,

$$\left(\mu - \frac{ic\kappa}{p}\right)^2 = 1 - \frac{p_0^2}{p^2 - ip\nu} \qquad \text{.....(6)},$$

or, separating real and imaginary parts,

$$\mu^2 - \frac{c^2\kappa^2}{p^2} = 1 - \frac{p_0^2 p^2}{p^4 + p^2\nu^2}$$

and $$2\mu\frac{c\kappa}{p} = \frac{p_0^2 p\nu}{p^4 + p^2\nu^2} \qquad \text{.....(7)}.$$

If now, as is usually the case, $\frac{c^2\kappa^2}{p^2} \ll \mu^2$ we have, with a little rearrangement,

$$\kappa = \frac{\nu}{2c}\left(\frac{1 - \mu^2}{\mu}\right) \qquad \text{.....(8)},$$

an important relation between κ and μ.

§ 3. THE RELATION BETWEEN EQUIVALENT PATH, OPTICAL PATH AND ELECTRIC WAVE FREQUENCY

The most fundamental quantity measured by radio methods is the group-time of flight of the waves on the journey to the ionosphere and back. The product of this quantity and the velocity of light is termed the equivalent path P' of the waves. We have, in this way,

$$P' = c\int \frac{ds}{U} \qquad \text{.....(9),}$$

where U is the group-velocity of the waves along an element of track ds. For the case of vertical incidence P' is evidently equal to twice the equivalent height h'. Since U is less than c it is also seen that the equivalent height is always greater than the actual height of reflection.

Now the equivalent path is related to the optical path P by the relation[13]

$$P' = P + f\frac{\partial P}{\partial f} \qquad \text{.....(10),}$$

where f is the electric wave frequency. We have therefore,

$$P = \frac{1}{f}\int_0^f P' \,.\, df \qquad \text{.....(11),}$$

which shows that, if the experimentally-determined relation between P' and f can be extrapolated to zero frequency, an estimate may be made of the magnitude of the optical path[15]. Now since

$$P = \int \mu \, ds \qquad \text{.....(12),}$$

and $\mu < 1$ we see that, for vertical incidence, the actual height reached by the waves is greater than half the optical path, and less than half the equivalent path. Thus

$$\frac{P'}{2} > \frac{s}{2} > \frac{P}{2} \qquad \text{.....(13).}$$

Further progress may be made by employing the approximations mentioned above for the case of the ordinary wave. Neglecting frictional effects, (6) becomes

$$\mu^2 = 1 - \frac{p_0^2}{p^2} \qquad \text{.....(14),}$$

from which it may be shown that

$$U = c\mu \qquad \text{.....(15),}$$

and

$$P' = \int \frac{ds}{\mu} \qquad \text{.....(16).}$$

In this special case the actual height $\frac{1}{2}s_0$ reached by waves of frequency f_0 may be estimated from the relation[15], [16]

$$\frac{s_0}{2} = \frac{1}{\pi}\int_0^{f_0} \frac{P'}{\sqrt{f_0^2 - f^2}}\, df \qquad \text{.....(17),}$$

Another simplification can be effected for these special conditions, for (8), with the aid of (12) and (16), becomes

$$\int \kappa \,.\, ds = \frac{\nu}{2c}[P' - P] \qquad \text{.....(18).}$$

Now $\int \kappa \,.\, ds$ is $-\log \rho$ where ρ is the reflection coefficient of the ionosphere, a measurable quantity. Further it has been shown[17] that, from (10) and (18), we have

$$\nu = -\frac{2c}{f}\frac{\partial (f \log \rho)}{\partial P'} \qquad \text{.....(19),}$$

so that ν, the frequency of electron collisions with the gas molecules may be estimated. It should be noted, however, that this equation is only applicable to cases where ν is approximately constant over the ray track where P' differs appreciably from P.

The discussion in the present section has so far been written in terms of a ray treatment. But it is well known that it is not possible adequately to describe optical phenomena in terms of geometrical optics when the refractive index μ varies appreciably within a wave-length (in the medium) in a direction normal to the wave front. Now for total reflection at normal incidence the wave-length becomes indefinitely large as $\mu \to 0$ so that in this case the ray treatment needs further justification. In three important papers Hartree[18] has discussed, from a wave standpoint, the reflection from a stratified medium. General expressions for the optical and equivalent paths in terms of the solutions of wave propagation are given, and exact expressions obtained for stratified media with certain specified variations of refractive index. On the whole, however, it may be said that the work of Hartree shows that the errors made in using a ray treatment are not serious in most practical problems.

§ 4. THE MEASURABLE QUANTITIES

No attempt can be made here to describe in detail all the ionospheric observations which have been made in different parts of the world. Instead, an attempt is made in this section to show the *kind* of information which can be obtained by radio methods, a brief index of such methods being given, while in the succeeding sections the results of the applications of these methods are used to give as complete a description of the nature of the ionosphere as seems possible at present.

(*a*) *Measurement of group-time of flight* (τ)

To measure the group-time τ for a signal to travel to the level of reflection and back to the ground it is necessary to impress some kind of distinguishing mark on the waves, so that it may be recognized on its return. Now an electric wave has frequency, amplitude and polarization and the first two are conveniently varied experimentally. Thus both frequency-modulation and amplitude-modulation methods are available for the measurement of τ.

(i) *Frequency-modulation method* (Appleton and Barnett[19]). Here the frequency of an emitter is varied continuously through a small range Δf and the number Δn

of the resulting interference maxima between the ground and atmospheric waves is noted. The difference between the group-times of travel for the ground and atmospheric waves is then equal to $\frac{\Delta n}{\Delta f}$ so that, since the time of transit of the ground wave is calculable (in vertical incidence experiments it is negligible), τ can be determined.

(ii) *Amplitude-modulation method* (Breit and Tuve[20]). Brief trains of radio waves (e.g. lasting 10^{-4} sec.) are emitted at suitable intervals and the ground and echo-waves recorded by a high-speed galvanometer or oscillograph with calibrated time-scale. The value of τ is thus measured directly. This method of measurement is by far the most generally useful because of the ease with which, by means of it, multiply-reflected echoes are recognized. It also lends itself admirably to automatic methods of registering the variation of τ either with time or with electric wave frequency.

(iii) *Combination method* (Appleton[21]). Here both amplitude- and frequency-modulations are simultaneously used. The method is specially useful for investigating the structure of magneto-ionic doublet echoes which, with either method (i) or (ii), are unresolved.

When τ is known, the value of P' (i.e. $c\tau$) the equivalent path can be calculated. Most of the wireless observations made in connexion with geophysical enquiry consist in determining the relation between P' and time t, or the relation between P' and frequency f. Automatic methods of recording the nature of both relationships have been developed. In the first case[22] such a record gives us a (P', t) curve for a certain specified frequency. In the second case[23] a (P', f) curve for a certain set of ionospheric conditions is obtained. The latter type of curve is of great value, for from it can be read off the critical penetration frequencies for the different ionospheric regions from which ionization magnitudes may be deduced. By far the greater part of the succeeding sections is concerned with the discussion of information derived from (P', f) records.

(*b*) *Measurement of temporal rate of change of optical path*

The magnitude of the rate of change of optical path $\frac{dP}{dt}$ of downcoming waves but not the sign of this quantity, may be measured by a modification of the group-retardation method proposed by Breit[24]. The magnitude and sign of $\frac{dP}{dt}$ may be measured by the frequency-change method[25]. Using the latter, it has been shown, for example, that in the early morning, when upper-atmospheric ionization is increasing, the value of P' falls at a greater rate than does the value of P. This has been interpreted as indicating that the region of reflection has not a sharp boundary; and that the fall of P' at sunrise is due to the increase of the ionization gradient of the reflecting region rather than to the fall of the region as a whole.

(*c*) *Measurement of angle of incidence of downcoming waves*

(i) *Stationary-emitter method* (Appleton and Barnett[26]). Here the value of the angle of incidence is measured from the ratio of the signals simultaneously produced by a downcoming wave in open aerial and loop aerial systems.

(ii) *Moving-emitter method* (Mirick and Hentschel[27]). Interference phenomena between direct and indirect waves are examined as the emitter moves towards or away from the receiving station.

When the angle of incidence of the reflected waves and the distance between the sending and receiving stations is known, the equivalent height of reflection can be obtained by simple triangulation.

(*d*) *Measurement of lateral deviation of atmospheric waves*

Very often the plane in which the waves travel from sender to receiver *via* the ionosphere does not coincide with the vertical plane in which these stations lie. The departure can be investigated by means of

(i) *The Adcock direction-finder* (Smith-Rose and Barfield[28]).

(ii) *The spaced-receiver method* (Ratcliffe and Pawsey[29]).

(iii) *The modified stationary-emitter method* (see (*c*) i) (Martyn and Green[30]).

The existence of the lateral deviation of waves in the process of reflection, first suspected by Parkinson[31] and by Cherry and Martyn[32], was reliably confirmed by Ratcliffe and Pawsey[29] in a very complete study of the characteristic features of downcoming waves. These last-mentioned authors conclude that for waves of 200 to 500 metres in length, and for distances of transmission inferior to 200 km., the reflected wave must be considered as the result of ionospheric "scattering" from a series of diffracting centres distributed over an area of radius about 20 km. Martyn and Green[30] have found that under somewhat similar conditions the measured angles of incidence approach the values corresponding to symmetrical reflection but that a considerable amount of lateral deviation of the downcoming ray is normally present.

(*e*) *Measurement of polarization of downcoming wireless waves*

(i) *The frequency-change method* (Appleton and Ratcliffe[33]). Interference phenomena are simultaneously recorded on loop aerials of different orientation and from the phases and amplitudes of the recorded interference "fringes" the polarization of the downcoming waves may be deduced.

(ii) *The cathode-ray polarimeter for pulse emissions* (Appleton and Watt[34]). Here the polarization of the echoes obtained from a station emitting pulses (see (*a*) ii) is delineated as rotating traces on a cathode-ray oscillograph screen.

(iii) *Use of aerial system responding to right or left circularly polarized downcoming waves* (Ratcliffe and E. L. C. White[35]; and Eckersley[36]). The method of Ratcliffe and E. L. C. White is of special importance because of its application in automatic recording of (P', t) relations, using the Breit and Tuve method of measuring P'.

(iv) *The cathode-ray polarimeter for continuous emissions* (Ratcliffe and F. W. G. White)[37]. Here the polarization of the downcoming waves from a distant emitter is portrayed as a rotating trace on a cathode-ray oscillograph.

Measurements of the polarization of downcoming waves have proved important in demonstrating the presence of free electrons in the ionosphere and in the general confirmation of the magneto-ionic theory from which are deduced the relevant formulae for the calculation of the ionization content of the ionosphere.

(*f*) *Measurement of reflection coefficient of the ionosphere*

By reflection coefficient is meant here the ratio of the emergent and incident rays as measured in the unionized region immediately beneath the ionosphere. Using the notation of section (3), we have

$$\log \rho = -\int \kappa \, . \, ds \qquad \ldots\ldots(20).$$

(i) *The use of multiply-reflected echoes* (Appleton[38]). Allowance is made for the fact that the doubly-reflected echo has made a double journey to the layer and back so that in the case of vertical incidence ρ is equal to twice the ratio of the second- and first-order components. Methods of the automatic registration of echo-amplitudes have been developed by Hollmann and Kreielsheimer[39] and by F. W. G. White[40].

Measurements of ρ have been used in the study of the absorption of wireless waves in the ionosphere and in the measurements of the collisional frequency of electrons with the gaseous molecules.

§ 5. THE IONOSPHERIC STRUCTURE

Almost all the early work on the radio exploration of the ionosphere consisted in the determination of the equivalent height of reflection for waves of a given fixed frequency. We may say that it provided us with (P', t) or (h', t) curves. It is now recognized that far more information is to be obtained from curves illustrating the (P', f) or (h', f) relations. Nevertheless certain deductions were made from early (P', t) curves which have been confirmed later. In 1926 from observations with 0·75 Mc./sec. waves it was deduced that the ionosphere consists of two main ionized regions which were termed Region E (Kennelly-Heaviside Layer) and Region F, the latter being the upper and more intensely-ionized region[41]. In 1931 Schafer and Goodall[42], in America, confirmed the two-region hypothesis. From (P', t) curves with 3·75 Mc./sec. waves Appleton and Builder[43], in 1931, observed the phenomenon of magneto-ionic splitting in the case of Region F echoes.

No attempt, however, is made here to describe fully the work done in connexion with (P', t) observations, since all the results obtained may be illustrated in the more general discussion of (P', f) determinations to which we now proceed.

In figure 2 is shown a typical (P', f) curve for conditions at noon on a spring day (April 10, 1933) in south-east England. The equivalent path P' is expressed in kilometres and the electric wave frequency in megacycles per second. Measurements made for the ordinary wave (which is of left-handed circular polarization)

are indicated with rings, and the values for the extra-ordinary wave (which is of right-handed polarization) are indicated by crosses. We consider the ordinary ray curve first. For frequencies below 3·1 Mc./sec. reflection takes place at Region *E*. For frequencies higher than this critical value, reflection takes place from Region *F*. Thus 3·1 Mc./sec. is the ordinary ray critical frequency for Region *E* and, using (4), the value of N_{max} for Region *E* is calculated as $1{\cdot}2 \times 10^5$ electrons per cm^3.

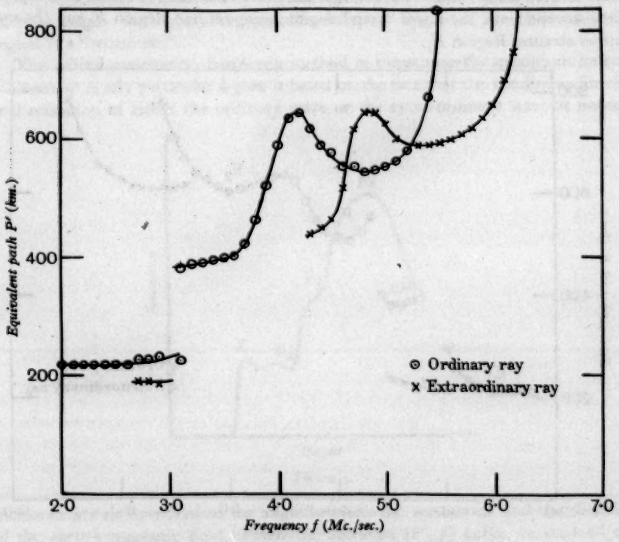


Figure 2. (P', f) curve for 1200 G.m.t., April 10, 1933.

Following the Region *F* curve, it is seen that for frequencies higher than 5·5 Mc./sec. the ordinary ray is not returned from Region *F*. This frequency is then taken as the penetration frequency for Region *F* so that in this case the value of N_{max} is correspondingly found to be $3{\cdot}75 \times 10^5$ electrons per cm^3. It will be seen that the curve for the extra-ordinary ray has quite a similar trend to that found for the ordinary ray, but is displaced about 0·67 Mc./sec. on the frequency scale. (Between 3·0 and 4·3 Mc./sec. the extra-ordinary ray echoes are missing, probably due to absorption.) This is due to the fact that, as follows from the magneto-ionic theory, the refractive index of a given ionized medium for an ordinary ray of frequency f Mc./sec. is approximately the same as for an extra-ordinary ray of frequency $(f + 0{\cdot}67)$ Mc./sec.

The maximum in the (P', f) curve for Region F indicates that the region cannot be a simple one, and the fact that for frequencies higher than 4·2 Mc./sec. (see ordinary ray curve of figure 2) the value of P' falls with increasing f indicates that there must be a sudden increase of ionization with height at a particular level. In other words the (P', f) curve indicates that Region F is composite, with the lower and less ionized component forming a "step" on the lower side of the higher and more ionized component[44]. When necessary it is customary to distinguish between these components, the lower "step" region being termed Region F_1 and the main upper stratum Region F_2.

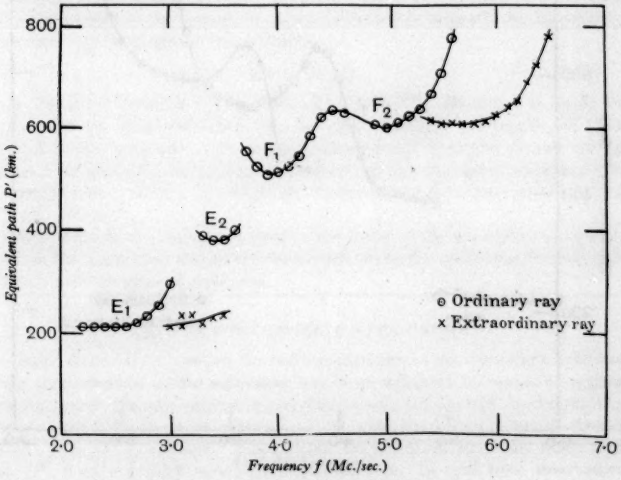


Figure 3. (P', f) curve at 1200 G.m.t., April 8, 1933.

Evidence is sometimes observed[44] of reflection from a region between Region E and Region F. This is termed the intermediate region. A (P', f) curve illustrating conditions when reflection was noted from the intermediate region is shown in figure 3. In this diagram the particular part of the (P', f) trace indicating reflection at each region is marked.

The radio evidence as a whole, therefore, indicates that the structure of the ionosphere in south-east England must be somewhat as indicated in figure 4 where the relative values of the ordinates are approximately correct but not those of the abscissae.

The "step" Region F_1 is only marked when the sun is up, for at night it is practically undetectable; Regions F_1 and F_2 merging into a simple region. It is

clear that any region higher than Region F_2, but of lower ionization, would not be detected by the usual wireless methods of exploration. The occurrence of reflections from levels between Regions E_1 and F_1 indicates that we must not regard Region E and F as layers with an unionized region between. A more correct, albeit approximate, description is that Region E (i.e. E_1 and E_2) constitutes an ionized shell of fairly constant ionization and about 100 km. thick, immediately underlying a much more intensely ionized region which has a maximum content at a certain height. In slightly less scientific language Region E is a "plateau" and Region F a "mountain".

The critical penetration frequency method of measuring the maximum ionization content in any particular region is based on the fact that the conditions for the total reflection of either the ordinary wave or the extra-ordinary wave at normal

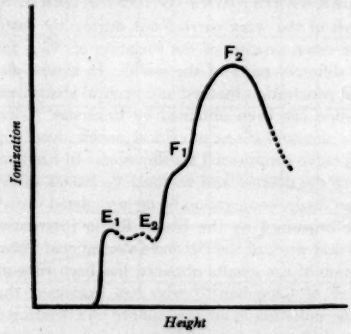


Figure 4.

incidence are independent of the angle between the wave-track and the direction of the earth's magnetic field. From the observed (P', f) curve we read off the critical penetration frequencies for both ordinary and extra-ordinary waves and use (4) and (5) to calculate the value of N_{max} in the region in question.

Since one formula is merely a check on the other, we usually specify only the ordinary ray critical frequencies of the regions E_1, E_2 (when detectable), F_1 (during the daytime) and F_2 in a range of increasing values. These are designated usually as $f_{E_1}{}^0$, $f_{E_2}{}^0$, $f_{F_1}{}^0$ and $f_{F_2}{}^0$ respectively, while the extra-ordinary ray critical frequencies are, correspondingly, written $f_{E_1}{}^X$, $f_{E_2}{}^X$, $f_{F_1}{}^X$ and $f_{F_2}{}^X$. Sometimes during the daytime a series of minor maxima are to be noted on the (P', f) curve for the frequency ranges below $f_{E_1}{}^0$ and between $f_{E_2}{}^0$ and $f_{F_1}{}^0$, but no systematic study of such phenomena has been published.

Various observations suggesting the occasional reflection of waves at levels higher than that of Region F_2 have been reported. For example Appleton and Builder[43], using 80-metre waves, recorded the reception of an echo from an equi-

valent height of 600 km. simultaneously with normal reflection from Region *F* to which it could not be related as a multiple reflection. Mimno[46], using 85-metre waves, found series of echoes also of 600 km. equivalent height together with another series from equivalent heights of 1100 to 1800 km. His observations led him to class the first series as reflected from a Region *G* and the second series as from a Region *H*. Kirby and Judson[47] have also confirmed the existence of echoes from an equivalent height of about 600 km. which are not to be related simply with Region F_2 echoes. With increasing frequency they find the equivalent height increases up to 1100 km. at 8·0 Mc./sec. It has not, however, been demonstrated with complete certainty that such echoes of medium delay time correspond to waves reflected at vertical incidence.

§6. REGULAR VARIATIONS OF IONOSPHERIC STRUCTURE

The greater part of the work carried out during the last few years has been concerned with the determination of the variation of N_{max} for the different ionospheric regions in different parts of the world. In almost all cases this has been done by the critical penetration method and vertical transmission, though valuable additional information has been obtained by Eckersley[48] for the special case of Region F_2 from the determinations of critical penetration frequencies observed in long-distance short-wave commercial transmission. In England and America fairly complete surveys of the diurnal and seasonal variations of N_{max} for Regions E_1, F_1 and F_2 have been made, cooperation being stimulated initially in this connexion by the programme organized by the Union Radio Internationale Scientifique as part of the geophysical work of the Second International Polar Year 1932–33.

The interpretation of the results obtained has been very materially assisted by the theoretical work of Chapman[49] who has examined the ionizing effect of monochromatic solar radiation in an atmosphere on a rotating earth. Millington[50] has further elucidated Chapman's work by preparing a number of charts, indicating the theoretical variations of ionization, in a form easily comparable with the experimental results. Chapman's theory is based on relatively simple assumptions. The radiation is considered as suffering absorption according to the simple mass-absorption law, while the recombination of ions and electrons is assumed to be the same at all levels of the atmosphere. We consider first the rate of production of ions per c.c. which is, of course, proportional to the rate of absorption. Suppose that the molecular density ρ of an ideal atmosphere varies with height h according to the formula

$$\rho = \rho_0 e^{-\frac{h}{H}} \qquad \text{.....(21),}$$

where H is a constant which is proportional to the absolute temperature T. Let I be the rate of absorption of the radiation by the atmosphere, at height h, when incident at the angle χ. Then it may be shown that

$$I = I_0 e^{1 - z - e^{-z}\sec\chi} \qquad \text{.....(22),}$$

where
$$z = \frac{h - h_0}{H}, \quad I_0 = \frac{S_\infty}{H \exp 1}, \quad h_0 = H \log A\rho_0 H,$$

A being the absorption coefficient and S_∞ the intensity of the radiation outside the earth's atmosphere ($h=\infty$). Thus z is the height, measured in terms of H as unit, reckoned from the level h_0. This is the level at which, when $\chi=0$ (i.e. for direct incidence, as at the equator at the equinox at noon) I attains its maximum value I_0. The density at height h_0 is $1/AH$.

The graph of I/I_0 as a function of height z is shown, for $\chi=0$ in figure 5. Since, as has been said, the rate of ion-production q is proportional to I, the curve illustrates also the variation of ion-production with height when the sun is vertically overhead and thus illustrates theoretically the formation of a simple ionized region. In the wireless experiments we are, however, concerned not with the rate of ion-

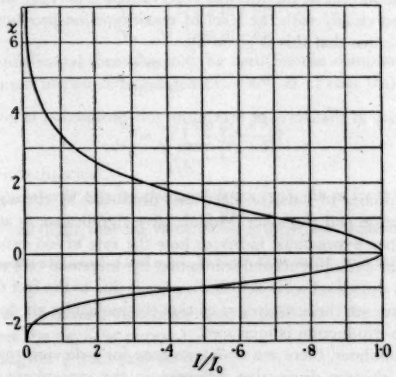


Figure 5.

production q, but with the number per cm^3, N, of either positive ions or electrons. These quantities are, in the simple case contemplated, related by the equation

$$\frac{dN}{dt}=q-\alpha N^2 \qquad \text{.....(23),}$$

where α is the coefficient of recombination and t the time. In Chapman's investigation the daily variation of N at different heights, seasons and latitudes has been determined, q being assumed to be proportional to I, which varies with height and the sun's zenith distance as indicated by (22). His results are plotted in a series of curves illustrating the variation of N with time and with height. These show that, when reasonable values for the parameters are assumed, the value of N_{max}, which is the quantity measured in the critical penetration frequency method, reaches a maximum at or slightly after noon, decreases rapidly as sunset is approached and continues to decay throughout the night until a short time before ground sunrise when the matutinal resumption of ion-production causes the value to rise rapidly and then more slowly until the daytime maximum is again reached. In a

general way this represents the actual variation of N_{max} which has been found experimentally. The agreement between theory and experiment is good for the cases of Region E_1 (Kennelly-Heaviside Layer) and for Region F_1, while for Region F_2 there are found important discrepancies between the simple theory and the experimental results.

Up to the present there has been no attempt to correlate the experimental results with Chapman's diurnal ionization curves, for it is found that the readiest check between theory and experiment is available in connexion with the seasonal variation of noon-ionization. In this case the relevant theoretical relations assume a very simple form, as is shown below.

Bearing in mind that in the wireless observations it is N_{max} which is measured, we are concerned chiefly with the level of maximum ion-production q_{max}. From (22) it may be shown that this is given by

$$q_{max} = \gamma S_\infty \frac{\cos \chi}{H \exp 1} \qquad \ldots\ldots(24),$$

and the air-density at that level of maximum ion-production is given by

$$\rho = \frac{\cos \chi}{AH} \qquad \ldots\ldots(25).$$

Here γ is the number of ions (or electrons) produced by the absorption of unit quantity of radiation and the other symbols have significance as stated above. The first of the last two expressions indicates how the rate of ion-production depends on the angle of the sun. The second shows that the increased rate of ion-production consequent on a decreased solar zenith distance is due to the fact that the radiation penetrates further into the atmosphere so that the molecular air density at the level of maximum ion-production is increased.

As will be seen later, there are strong reasons for believing that recombination is the principal electron-dissipative influence in the ionosphere, so that we may write, in addition,

$$\frac{dN}{dt} = q - \alpha N^2 \qquad \ldots\ldots(23).$$

Moreover, if we are concerned with a noon temporal maximum of ionization for any particular region, we can write $\frac{dN}{dt} = 0$ and thus

$$q = \alpha N^2 \qquad \ldots\ldots(26).$$

This equation applies to the level of spacial maximum of ionization since, α being constant, the level of maximum ionization coincides with the level of maximum ion-production.

If now we wish to compare the values of maximum N for noon in summer and winter we have

$$\frac{N_s}{N_w} = \sqrt{\left(\frac{q_s \alpha_w}{q_w \alpha_s}\right)} \qquad \ldots\ldots(27),$$

where subscripts have been used to distinguish between the summer and winter

values. Also from (24), we have, bearing in mind that H is proportional to the absolute molecular temperature,

$$\frac{q_s}{q_w}=\frac{\cos\chi_s}{\cos\chi_w}\cdot\frac{T_w}{T_s}=\frac{\sin(\theta+\delta)}{\sin(\theta-\delta)}\cdot\frac{T_w}{T_s} \qquad \text{.....(28),}$$

where T is the molecular temperature, θ is the co-latitude of the place of observation and δ is the sun's maximum declination. We have, therefore, from (27) and (28),

$$\frac{N_s}{N_w}=\sqrt{\left(\frac{\sin(\theta+\delta)}{\sin(\theta-\delta)}\cdot\frac{\alpha_w}{\alpha_s}\frac{T_w}{T_s}\right)} \qquad \text{.....(29),}$$

or, assuming that α and T do not vary from summer to winter

$$\frac{N_s}{N_w}=\sqrt{\left(\frac{\sin(\theta+\delta)}{\sin(\theta-\delta)}\right)} \qquad \text{.....(30).}$$

The same fundamental formulae may be used for the comparison of the noon maximum ionization at two different places A and B. From (24) we have corresponding to (30)

$$\frac{N_A}{N_B}=\sqrt{\left(\frac{\sin(\theta_A+\delta)}{\sin(\theta_B+\delta)}\right)} \qquad \text{.....(31),}$$

δ being the sun's declination.

Another simple result can be deduced from (23). It is found, for Regions E and F_1 during full daylight, that the value of $\frac{dN}{dt}$ is in general small numerically compared with estimated values of q and αN^2. We can thus write, to a first approximation

$$q=\alpha N^2 \qquad \text{.....(26).}$$

(This equation is, of course, rigorously true at mid-day. We are here assuming that it is true over the daylight period.)

On substituting for q from (24) we have

$$N=\text{const.}\ \sqrt{\cos\chi} \qquad \text{.....(32),}$$

a relation expressing the dependence of N on the angle of the sun. Also since, according to (4), the value of N is proportional to the square of the ordinary ray critical penetration frequency f^0, this becomes

$$f^0=\text{const.}\ (\cos\chi)^{\frac{1}{4}} \qquad \text{.....(33).}$$

(*a*) *Experimental results in temperate latitudes: south-east England*

The first experimental determinations of the diurnal and seasonal variation of critical penetration frequencies were made during 1931 by Appleton and Naismith[31]. These were concerned with Region E and showed that the noon equinox ordinary ray critical frequency was about 3 Mc./sec. indicating a maximum ionization content of about $1{\cdot}1\times10^5$ electrons per cm^3. For the same season the midnight value of $f_E{}^0$ was about 0·8 Mc./sec. indicating a maximum ionization content of $7{\cdot}9\times10^3$ electrons per cm^3. For all days the ionization in Region E was found to reach a maximum at noon, falling off steadily as sunset approached. The ionization content was found to fall to a low value during the dark hours, a minimum being

reached just before ground sunrise. Somewhat similar measurements were made by Appleton and Builder for Region F who found that the daytime ionization in that region was about 4 or 5 times[51] as intense as that in Region E. Evidence of a nocturnal increase in maximum ionization content for Region F was also noted[52].

The very early measurements of critical penetration frequencies were not very accurate, largely because the experimental technique then available did not permit the rapid determination of the equivalent path P' for a large number of values of f round the critical value. For example, in the early work a frequency spacing of 0·25 Mc./sec. had to be adopted[51]. This was reduced by Appleton to a spacing of 0·05 to 0·1 Mc./sec. until ultimately continuous registration[23] of the relation between P' and f was effected by Gilliland, Naismith and Bainbridge-Bell. The contributions to the technique of this subject made by the last-mentioned three

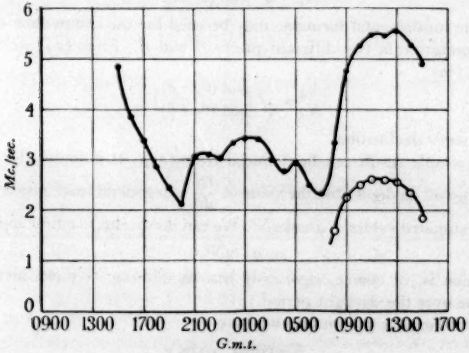


Figure 6. Winter: ● $f_{F_2}{}^{\circ}$; ⊙ $f_{E_1}{}^{\circ}$.

authors constitute a great advance and now critical penetration frequencies can be determined with considerable accuracy. The influence of this has been that the observations made as the work has progressed since 1931 have increased steadily in accuracy.

Examples of the measurements[53] made by determining critical penetration frequencies from a continuous (P', f) record are exhibited in figures 6, 7 and 8.

In figure 6 are exhibited the ordinary ray critical frequencies for Region F_2 for a winter day. The day values for Region E are also included for comparison. It will be seen that the maximum critical frequency for Region F_2, as also for Region E, is reached at noon. It is also seen that the maximum ionization content in Region F_2 increases during the dark hours. This is characteristic of winter conditions. In figure 7 are shown corresponding values for a summer day. In this case the Region F_1 critical frequencies, which can be measured in summer, are also included. Here it is seen that while the critical frequencies (which are, of course, propor-

tional to $\sqrt{N}$) have a symmetrical trend about noon for Regions E_1 and F_1, the same cannot be said for Region F_2. In the latter case there is a pre-noon maximum of ionization, a minimum shortly after noon, and a pronounced afternoon maximum about 1900 to 2000 G.m.t. There is in this case no nocturnal increase of ionization as is noted in the winter months.

In figure 8 are exhibited the noon values for the ordinary ray critical frequencies for Regions E, F_1 and F_2 for the period December 1st, 1933, to November 30th, 1934. It will be noticed that both Regions E and F_1 exhibit a definite seasonal variation, the values being higher in summer than in winter, whereas for Region F_2 there is no such corresponding variation. The average summer noon value for Region E

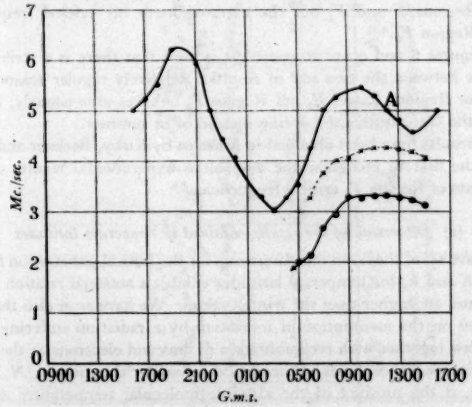


Figure 7. Summer: ●—● $f_{F_2}°$; ×- - -× $f_{F_1}°$; ⊙—⊙ $f_{E_1}°$.

ordinary ray critical frequency is 3·3 Mc./sec. and the corresponding winter value is 2·45 Mc./sec. This means that the ratio of summer to winter noon ionization is $(3{\cdot}3/2{\cdot}45)^2$ or 1·8. The middle series of values in figure 8 refers to Region F_1 which is only distinguishable in the daytime. From the average summer and winter critical frequency values we estimate the ratio of summer to winter noon ionization for Region F_1 to be 1·86.

It will be seen that the trend of the ordinary ray critical-frequency curve for Region F_2, which is the densest of the normally occurring regions of the ionosphere, differs markedly from that for the two lower regions E and F_1. Not only are there marked differences in the day-to-day values, but, instead of the expected summer maximum, we note that the values at that period are relatively low. There are two rather ill-defined maxima, one in February-March and the other in October-November, the latter being the greater.

(b) Experimental results in temperate latitudes continued: America

Extensive valuable surveys of ionospheric conditions in America have been made by Schafer and Goodall[54]; by Gilliland[55]; by Kirby, Berkner and Stuart[56] and by Kirby and Judson[57]. In all cases the critical frequency method of ionization measurement was adopted. There is general agreement that the ionospheric structure in America differs very little from that found in south-east England, so that a comparison of the results obtained in the two latitudes can be made. As an example of the American results figure 9 shows the noon critical frequencies for Regions E, F_1 and F_2 at Deal, New Jersey (Lat. 40° 15′ N.) as measured by Schafer and Goodall. Here it may be noted that the ordinary ray critical frequencies are given for Regions E and F_1 but the extra-ordinary ray critical frequencies are shown for Region F_2.*

When figures 8 and 9 are compared it is seen that there is a striking general resemblance between the two sets of results. Relatively regular seasonal changes are noted for Regions E and F_1 but Region F_2 behaves anomalously, there being maxima in the late autumn and spring instead of in summer.

Similar results have been obtained in America by Kirby, Berkner and Stuart[56], who were the first to recognize the anomalous experimental results obtained in measurements of Region F_2 critical frequencies.

(c) Discussion of the results obtained in temperate latitudes

As we have seen, the noon measurements for the critical penetration frequencies of Regions E and F_1 for temperate latitudes exhibit a seasonal relation in that the summer values are higher than the winter values. We have seen also that a simple theory, based on the assumption of ionization by a radiation suffering the mass-absorption law together with recombination of ions and electrons as the process of ionization disappearance, leads to theoretical values for the ratio N_s/N_w. We have, for example, if the product of the absolute molecular temperature and the recombination coefficient is assumed to be the same in summer and winter,

$$\frac{N_s}{N_w} = \sqrt{\left(\frac{\sin(\theta+\delta)}{\sin(\theta-\delta)}\right)} \qquad \text{.....(34).}$$

In the subjoined table is shown a comparison of the theoretical and experimental values of this ratio, the latter being the values given in the references cited.

Place	Latitude	Reference	Region	N_s/N_w experimental	N_s/M_w calculated
Slough	51½° N.	(53)	E_1	1·8	1·84
Slough	51½° N.	(53)	F_1	1·86	1·84
Deal (U.S.A.)	40° 15′ N.	(54)	E_1	1·95	1·47
Deal (U.S.A.)	40° 15′ N.	(54)	F_1	1·22	1·47
Washington (U.S.A.)	38° 55′ N.	(56)	E_1	1·3	1·44
Washington (U.S.A.)	38° 55′ N.	(56)	F_1	1·4	1·44

* Note that $f_{F_2}{}^o$ can be estimated in any instance since it is known that $f_{F_2}{}^X - f_{F_2}{}^o = 0{\cdot}75$ Mc./sec.

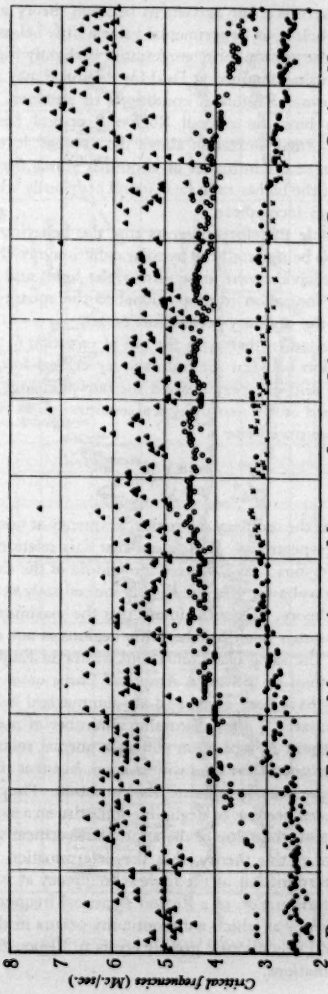


Figure 8. ● and × critical frequencies for E^s, Mc./sec.; ○ critical frequencies for F_1^s, Mc./sec.; ▲ critical frequencies for F_1^s, Mc./sec.

With one exception there is a fair agreement between theory and experiment. Moreover, as will be seen below, an experimental value a little below the theoretical value is to be expected if the atmospheric temperature is slightly higher in summer than in winter. The high value obtained at Deal for Region E may possibly be due to the influence of abnormal Region E conditions in summer. Appleton and Naismith[53] have shown how the normal Region E critical frequency can be measured even when abnormal ionization above the 100 km. level is found. By their being able to recognize the influence of abnormal effects they found a value of 1·8 for N_s/N_w instead of the higher ratio (2·2) found previously when the elimination of abnormal values was incomplete.

The evidence as a whole therefore suggests that the behaviour of Regions E and F_1 is normal, ionization being produced by solar radiation travelling rectilinearly (which other experiments have shown to be ultra-violet light) and that the rate of the disappearance of the ionization is proportional to the square of the concentration as is expressed in the ordinary recombination law.

The matter may be tested further with the aid of equation (33) in section (6) which expresses the relation between the ordinary ray critical frequency with the angle of the sun over the mid-day period when the rate of change of ionization is small. This relation is based on the same physical assumptions as the theory tested in the table. From (33) we may write

$$\frac{f^0}{f_N{}^0} = \left(\frac{\cos\chi}{\cos\chi_N}\right)^{\frac{1}{2}} \qquad \text{.....(35),}$$

where $f_N{}^0$ and χ_N represent the ordinary ray critical frequency at noon and the sun's zenith distance at noon respectively. It is found that this relation represents the variation of f^0 for both Regions E and F_1 over the middle of the day very closely.

Turning now to the results for Region F_2 it is immediately seen that there is no such agreement with theory. Here it is found that the maximum of the critical frequency observed at summer noon for the lower regions is not reproduced, the value of $f_{F_2}{}^0$ being about the same in summer and winter in England, and being actually less in summer than in winter in America. Three entirely different explanations of the anomaly have been given and are summarized below.

Kirby, Berkner and Stuart[56], after discussing a number of possibilities, came to the conclusion that Region F_2 ionization suffers a normal solar control, being more intense on a summer noon than on a winter noon, but that the critical penetration frequency technique fails to indicate this variation. They conclude that, due to absorption, the echoes become of negligible intensity on a summer noon at a frequency considerably below the value of the appropriate critical value. It should be mentioned, in support of this theory, that the determination of a Region F_2 critical frequency is a determination of the highest frequency at which echoes are obtained, whereas the determination of a Region E critical frequency involves the measurement of the frequency at which a discontinuity occurs in the (P', f) curve. Absorption effects are thus likely to cause greater errors in Region F_2 measurements than in Region E determinations.

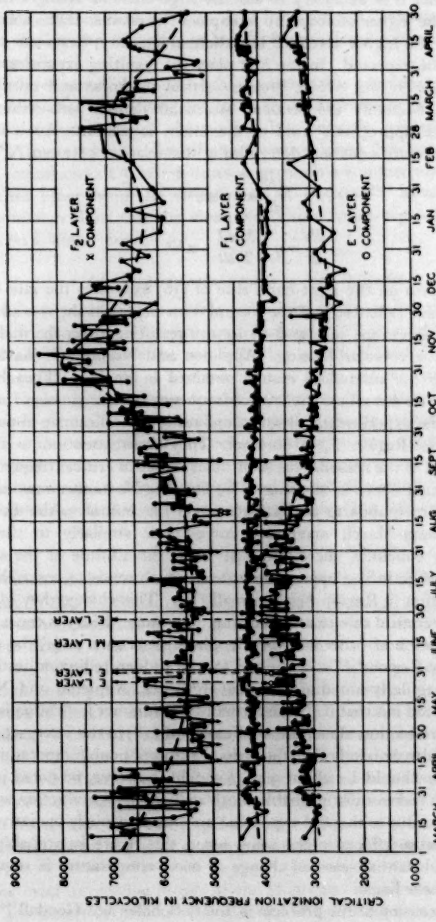

CRITICAL IONIZATION FREQUENCY IN KILOCYCLES
9000
8000
7000
6000
5000
4000
3000
2000
1000
E LAYER
M LAYER
E LAYER
E LAYER
F_2 LAYER
x COMPONENT
F_1 LAYER
O COMPONENT
E LAYER
O COMPONENT
15 31 15 30 15 31 15 31 15 31 15 30 15 31 15 31 15 28 15 31 15 30
MARCH APRIL MAY JUNE JULY AUG SEPT OCT NOV DEC JAN FEB MARCH APRIL
1933
1934

Figure 9.

Another explanation of the Region F_2 anomaly has been given by Appleton[58] who considers that it is necessary to assume large seasonal changes of molecular temperature in the higher atmosphere to explain the results. If the solar radiation, passing through the higher levels of low thermal capacity, heats the atmosphere, two effects may be expected. In the first place the resulting expansion will spread the upper atmosphere out over a greater vertical thickness and thus reduce the value of N_{max}. Secondly the reduced molecular density will reduce the ion-production at the appropriate level of maximum ionization. According to this theory the appropriate equation determining the relation between N, q, α and t is not

$$\frac{dN}{dt} = q - \alpha N^2 \qquad \text{.....(23),}$$

but

$$\frac{dN}{dt} = q - \frac{N}{T}\frac{dT}{dt} - \alpha N^2 \qquad \text{.....(36),}$$

where the first term on the right-hand side of (36) expresses the rate of ion-production by the solar radiation and the second term expresses the rate of dilution or concentration of electrons, ions and molecules resulting from thermal expansion or contraction due to solar heating. Appleton and Naismith[53] have used this theory to explain the anomalous results obtained in England. They believe that absorption effects do not vitiate the critical frequency measurements to any material extent and that the experimental observations really do indicate an abnormally low value for N_{max} in Region F_2 in summer. They draw attention to the analogy between the trend of the seasonal curve of noon values of critical frequency and the trend of the diurnal curve in summer. The October-November maximum on the seasonal curve corresponds to the late-afternoon maximum on the diurnal curve, while the February-March maximum corresponds similarly to the pre-noon maximum. They conclude, therefore, that when the altitude of the sun is high there is some change in the upper atmosphere which tends to make the resulting maximum ionization in Region F_2 abnormally low. This change they identify with a change of temperature due to solar heating. The same theory is considered[58] as giving an explanation of other ionospheric phenomena such as (*a*) the subdivision of Region F into Regions F_1 and F_2, and (*b*) the sudden reduction in the value of dN/dt which is regularly noted after sunset in winter. Appleton and Naismith[53] have further pointed out that if the abnormally low summer ionization is attributed to reduced ion-production alone, the application of (28) in the interpretation of the experimental results demands that the ratio of summer noon temperature to winter noon temperature should be about 3·4. It should, however, be noted in this connection that if, as seems most probable, both reduced ion-production and dilution of electron content due to thermal expansion are simultaneously operative and since both processes cause effects in the same sense, this figure is probably too high. Nevertheless a substantial seasonal change of noon temperature is required in an explanation on these lines.

The third discussion of the problem is due to Schafer and Goodall[54] who have drawn attention to the extreme variability of the day-to-day values of the Region F_2

critical frequencies. They, like Appleton and Naismith, accept the measured critical frequencies as reliable and suggest that, while the principal ionizing agency of the E and F_1 regions is ultra-violet light from the sun, "an important agency or mechanism causing the ionization in the F_2 layer is from some other more or less random source".

It might be expected that some help in solving this problem would be obtained from the information concerning critical penetration frequencies obtained for noon long-distance transmission of short-waves but here the evidence is conflicting. Hulburt[59], from a consideration of short-wave skipped distances, has shown that Region F_2 ionization is 1·5 to 1·8 times as great on a summer noon as on a winter noon. On the other hand the observations of Burrows[60] in transatlantic short-wave transmission show that the measured critical frequency is less on a summer noon than on a winter noon.

(*d*) *Experimental results in high latitudes*

A brief description of ionospheric conditions in Tromsø (Lat. 69° 40′ N.) has been given by Appleton, Naismith and Builder[61]. For conditions undisturbed magnetically they find there the two reflecting regions E and F. Maximum noon equinox ionization is found at Tromsø to be 60 to 70 per cent of that measured in south-east England. A striking feature of the results obtained at Tromsø was the absence of echoes during and after times of magnetic activity. Wagner[62] has also reported the disappearance of echoes during auroral displays and magnetic storms. Signals from distant stations were also found to be weakened. Similar results have been reported by Fleming[63] and Bontch-Bruewitch[64] for observations made at Fairbanks, Alaska, and Moormansk (Russia) (Lat. 68° 56′ N.) respectively.

(*e*) *Experimental results in low latitudes*

Observations in low latitudes have been made by Ranzi[65], Wells[68] and by Berkner and Wells[66]. Ranzi, working at Asmara (Lat. 15° 20′ N.), found there a late afternoon maximum of Region F ionization at 1800 h. local time. During the night, about 0200 h. local time, another maximum of critical frequency was noted. Region E was found always to be less strongly ionized than Region F and it also was found to be subject to nightly increases of ionic density.

The work of Berkner and Wells[66] confirms in part and extends in many ways that of Ranzi. Their observations were made at Huancayo (Lat. 12° 3′ S.). Region F_1 was found to show normal characteristics, the maximum value of ionization being reached at noon. On a February day, measurements made by the critical frequency method indicated that N_{max} in Region F_1 reached $2{\cdot}5 \times 10^5$ electrons per c.c. at noon. This may be compared with a value of $1{\cdot}9 \times 10^5$ electrons per c.c. which is measured for the same quantity in south-east England at the same time of the year.

But the most interesting results of the Huancayo measurements are concerned with the anomalous behaviour of Region F_2. They found a morning maximum of critical frequency at 0800 h. local time, a pronounced depression round about

noon and a very marked afternoon maximum at 1600 h. local time. The diurnal trend of critical frequency is thus somewhat similar to that noted in temperate latitudes on a summer day, except that in the latter case the maxima occur 2 to 3 hours later in local time. It should be added that Berkner and Wells report that the diurnal trend of critical frequency described does not always occur. Occasionally the morning maximum is greater than the afternoon maximum and sometimes the noon minimum is entirely replaced by a single maximum for the day. It does, however, appear that the pronounced mid-day depression of the critical frequency observed at Huancayo is the same as the phenomenon encountered in connexion with the low summer values obtained for the same quantity in temperate latitudes. The same theories of the effect must therefore be considered as possible explanations[53]. Hulburt[67], in a discussion of the Huancayo results, accepts the measured values as true critical penetration frequencies and explains the mid-day depression of maximum ionization as the result of atmospheric heating.

Measurements of the equivalent height of reflection for radio waves have been made by Rakshit[60] and by Mitra and Rackshit[70] in Calcutta (Lat. 22° 50′ N.). The former found that Region *E* reflections occurred at a height of 69 to 80 km. as compared with the 100 km. observed in temperature latitudes.

Wells[68] has made a study of the nature and characteristics of Region *E* reflections at Huancayo. He found that no marked critical frequencies are in evidence and that the transition of reflection from Region *E* to reflection from an upper ionized region is a gradual process. There is found, however, a definite and smooth diurnal variation of maximum ionization around noon.

§ 7. THE NATURE OF THE IONIZATION

At first it was thought that the reflection of waves by the ionosphere was caused by the influence of ions of molecular mass. The development of the magneto-ionic theory showed that detectable polarization effects should be evident if electrons were the effective electrical carriers. Appleton and Ratcliffe[33] determined the polarization characteristics of downcoming waves of 400 metres wave-length and found them to be approximately circularly polarized with a left-handed rotational sense. They explained this result as being due to differential absorption, the right-handed component having suffered almost complete extinction, as the magneto-ionic theory indicated should be the case for waves of that length. That an explanation on these lines was correct was demonstrated by Green[71] who found that, under comparable conditions, downcoming waves in Australia are approximately circularly polarized with a right-handed sense of rotation. These experiments therefore indicated that, for Region *E*, some electrons at least were present.

The discovery of doubling of Region *F* echoes[43] and the satisfactory explanation of all the observed phenomena in terms of the magneto-ionic theory leaves little doubt that only the electrons in this region affect radio waves appreciably. The fact that the ordinary component is of left-handed rotation in the northern hemisphere may be shown to indicate that the effective electrical carrier possesses a negative charge. The measured separation of the observed critical penetration frequencies

can be used to determine the ratio of the charge to the mass of the carriers if the value of the earth's total magnetic force at the level of reflection is assumed. The value determined in this way agrees closely with the known value for electrons. These matters are strikingly illustrated by the fact that any reflection effect due to electrons is dually represented on the (P', f) trace with a definite frequency spacing. For quasi-transverse transmission[13] this frequency spacing is, in south-east England, about 0·66 Mc./sec. and for quasi-longitudinal transmission 1·32 Mc./sec. For the portion of the (P', f) trace representing Region F reflections the dual trace with 0·66 Mc./sec. spacing is always found indicating that electrons cause the reflection of the waves.

In the case of Region E there is not correspondingly conclusive evidence. Polarization effects show the presence of some electrons, while magneto-ionic splitting has been recorded by F. W. G. White[72] and by Schafer and Goodall[54]. A dual representation of Region E effects on the (P', f) curve with a spacing of about 0·70 Mc./sec. has been reported by Appleton[73] and also by Appleton and Naismith[53]. On the other hand Kirby and Judson[57] in America find no evidence at all of magneto-ionic effects and, from this and other considerations, conclude that the normal Region E is preponderantly made up of heavy ions rather than electrons.

§ 8. THE NATURE OF THE NORMAL IONIZING AGENCY

One of the most striking features of the regular ionospheric variation is the profound solar control. The general trend of the diurnal variation curves of ionization and their marked correlation with sunset and sunrise leave little doubt that the major agency responsible for the ionization is of solar origin and travels in straight lines. It must therefore consist of either aetherial radiation (ultra-violet light) or neutral particles. A method of distinguishing between these alternative sources of normal ionization by means of radio observations during a solar eclipse has been proposed by Appleton and Chapman[74]. The test depends essentially on the difference between the speeds with which the two kinds of radiation travel towards the earth and the ratio of these speeds to the moon's orbital motion.

A consideration of the geometrical and time relations in the passage across the earth of the regions shaded by the moon from the sun's light and from the slower moving corpuscles shows that the corresponding eclipses will occur at different places and different times. They may, therefore, be distinguished by separate names; i.e., as the "optical" eclipse (the same for ultra-violet light as for visible light) and the "corpuscular" eclipse. It has been shown by Chapman[75] that, for corpuscles travelling at 1000 miles a second, the corpuscular eclipse will precede the optical eclipse by about 2 hours in time and will occur much to the east of the track of optical totality by a distance exceeding 1000 miles. If therefore ultra-violet light is the main ionizing agent in a particular region of the upper atmosphere, the electron density at a point in this layer above the track of optical totality should attain its minimum a few minutes after the central instant of optical totality; but if corpuscular ionization is the main cause, the minimum electron density should occur earlier, and to the east of the total shadow track.

Radio observations which include measurements[76, 77] of equivalent height of reflection are available for the eclipses of

(1) June 29th, 1927 (total) in England,
(2) August 31st, 1932 (total) in America,
(3) August 21st, 1933 (annular) in India,
(4) February 14th, 1934 (total) in the South Pacific (Losap Island),
(5) February 3rd, 1935 (partial) in America.

For eclipse (1) it was found that there was a partial return to night-time conditions, the maximum effect for Region E being observed round about totality. This result has been interpreted as indicating that ultra-violet light is the predominating ionizing influence for Region E. In the later eclipses the critical penetration frequency method of measuring maximum ionization content was adopted, so that a quantitative study of the ionization values was possible.

The experimental results in eclipse (2) show very definitely that ultra-violet light is to be regarded as the principal ionizing agency for Region E, for a marked diminution of ionization was recorded round optical totality. The minimum ionization values reached are given by Canadian and American observers as 42 and 30 per cent respectively. Chapman and Miller have determined theoretically the expected variation of ionization during the eclipse in Canada, assuming ultra-violet light as the sole ionizing agent and assuming the degree of recombination compatible with Appleton and Naismith's daily variation curves for Region E. The results agree closely with Henderson's[77] eclipse graph of ionization and show that there is no need to attribute any part of the residual 40 per cent electron content at totality to another ionizing source.

The eclipse observations also show that ultra-violet light is the principal ionizing agency for the "step" region F_1 since the observations in Nova Scotia[77] also indicated a minimum value of 40 per cent of normal coinciding approximately with optical totality. It may safely be said that the calculations of Chapman and Miller apply here similarly so that it may be concluded that there is no need to consider any ionizing agency other than ultra-violet light for Region F_1.

The work on Region F_2 does not lead to correspondingly definite conclusions. Observations at Kingston[77] were interpreted as indicating a reduction of 30 per cent from normal ionization at totality. On the other hand, the corresponding observations at Washington[77] did not show any optical eclipse effect, though the observers are of opinion that, due to absorption effects which cause region F_2 critical frequency measurements to be unreliable, the results did not preclude its occurrence. Observations at Deal, New Jersey, U.S.A.[77] also showed no changes of ionization which could be identified as an eclipse effect.

With regard to the effects of a corpuscular eclipse Region F_2 need alone be considered. Here the evidence is also conflicting. Fairly definite evidence of an effect is claimed by observers at Conway, New Hampshire and at Cambridge, Massachusetts[77], but other observers obtained no effect.

In the case of eclipse (3) evidence of an ultra-violet light control of Region E

ionization was obtained at Calcutta[77]. An anomalous effect which might have been corpuscular in origin was obtained for Region F_2.

At Losap Island[77] during eclipse (4) conclusive evidence of an ultra-violet light effect on both Regions E and F was obtained, but the conclusions drawn by different writers concerning the effects of a corpuscular eclipse are conflicting.

Additional evidence was obtained by Schafer and Goodall[76] during eclipse (5), who found that there was a decrease in the maximum ionization density of 20 to 25 per cent in Regions E_1, E_2 (Intermediate Region) and F_2 and that the minimum of ionization occurred at or shortly after totality. The percentage decrease was found to be progressively greater the higher the region, being approximately 20 per cent for Region E_1, 22 per cent for the Intermediate Region, and 25 per cent for Region F_2.

Taking the eclipse evidence as a whole it may therefore be regarded as certain that ultra-violet light is an important ionizing agency for all regions of the ionosphere, E_1, E_2, F_1 and F_2, that for Regions E_1 and F_1 it is the most important influence and that the possible existence of corpuscular effects in Region F_2 is still *sub judice*.

§ 9. THE PROCESS OF IONIZATION DISSIPATION

In interpreting the wireless evidence it is well to remember that the effect of ionization in producing reflection, refraction and absorption of the waves is measured, not by N the ionization content, but by N/m where m is the mass of one of the charged particles. In the case of the measurement of ionization by the ordinary ray critical frequencies we therefore estimate $\left(\frac{N_e}{m_e}+\frac{N_i}{m_i}\right)$, where the e subscripts refer to electrons and the i subscripts refer to ions. That N_e/m_e is greater than N_i/m_i is clearly proved for Region F_1 and F_2 by the magneto-ionic evidence. Some workers believe that the same can be said for Region E but agreement is not as general as is the case for Regions F_1 and F_2.

During both day and night the process of electron capture continues. At night when the solar influence is removed it is possible to study this process alone. This is most easily done experimentally in the case of Region F_2. There are two possible physical processes to be considered. Recombination of ions and electrons may take place, in which case we should expect the law

$$\frac{dN}{dt}=-\alpha N^2 \qquad \text{.....(37),}$$

or

$$\frac{1}{N}-\frac{1}{N_0}=\alpha t \qquad \text{.....(38),}$$

to hold at night, N_0 being the value of N when $t=0$.

Another possibility is that electrons may attach themselves to neutral atoms and, thus acquiring greater immobility, cease to be effective in causing radio wave reflection. In such case we should expect the law of electron-disappearance to be

$$\frac{dN}{dt}=-\beta N \qquad \text{.....(39),}$$

or

$$\log N_0-\log N=\beta t \qquad \text{.....(40).}$$

Eckersley[78], from a study of transatlantic facsimile transmission, has concluded that the recombination law expressed by (38) is obeyed and finds values for α ranging from $4{\cdot}25\times10^{-11}$ to $9{\cdot}5\times10^{-11}$. On the other hand, Appleton[79], using the vertical incidence method, does not find a linear relation between N^{-1} and t, the curve in question being concave upward indicating an increase of effective α (calculated from the slope of the curve expressing N^{-1} as a function of t) during a single night. For an April night, the variation of measured ionization being as shown in figure 10, the effective value of α was found to increase from $3{\cdot}7\times10^{-10}$ to $6{\cdot}0\times10^{-10}$, quantities which are nearly 10 times as large as those found by Eckersley.

It is true that Eckersley states that the accuracy of his measurements is not such as would allow the deviation (if any) from the recombination law to be measured and that only a 15 per cent accuracy can be expected. This is probably due to the

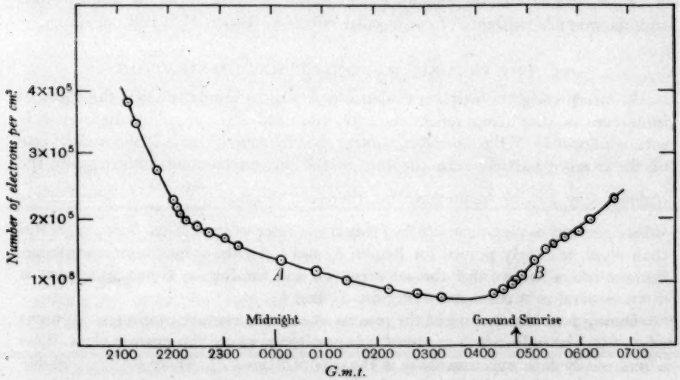


Figure 10. Curve showing variation of ionization in upper region, April 25–26, 1933.

fact that the value of N is measured by the facsimile method at different points along the transmission route at different times of the night. It therefore seems justifiable to accept the local ionospheric measurements as being of greater accuracy and we are therefore faced with the fact that the plot of N^{-1} with t is not linear. Since it has been found that the seasonal variation of N_{max} at noon for Regions E and F_1 are in agreement with the results predicted for a recombination theory as opposed to a theory of electron attachment, we should *a fortiori* expect recombination to be the predominant process of electron-dissipation also in Region F_2 where the air pressure, and thus the number of uncharged molecules per c.c. available for electron adherence, is less than at the lower levels. It is possible that the variation of effective α, as measured by the slope of the $\{N^{-1}, t\}$ curve, is due to the slow cooling and contraction of Region F_2 at night, the evidence for which

has already been noted above. In such a case the relevant equation describing the variation of N with t would not be (37) but

$$\frac{dN}{dt} = -\frac{N}{T}\frac{dT}{dt} - \alpha N^2 \qquad \text{.....(41).}$$

But even without a correction for temperature variation it is found that there is fair agreement between the values of α determined experimentally and the theoretical value for the same quantities calculated by Chapman from Milne's theory. For recombination of electrons and atomic oxygen ions Chapman[80] calculated that α at a temperature of 400° K. should be $3{\cdot}2 \times 10^{-10}$. For a temperature of 800° K., α, varying as $T^{-\frac{1}{2}}$, would be $2{\cdot}3 \times 10^{-10}$. These values are not far removed from the experimentally obtained magnitudes.

§ 10. PERIODICITIES, RECURRENCE TENDENCIES AND IRREGULARITIES

It has long been suspected that there exists a correlation between solar activity and upper-atmospheric ionization, both phenomena having a periodicity of a little over 11 years. Chapman[81], for example, has estimated from terrestrial magnetism considerations that the total conductivity of the ionosphere is 1·5 times as great at sunspot maximum as at sunspot minimum. The first indication of such a periodicity in the specific, as opposed to the total, conductivity of the ionosphere was provided by the experiment of Austin[82] who, from 1915 onwards, made measurements of the signal strength of 13,000-m. waves, emitted by Nauen and received in Washington. Austin found that there was a direct correlation between the signal strength and solar activity and, in conjunction with Wymore[83], showed that the signal strength was 1·8 times as intense at sunspot maximum as at sunspot minimum. Appleton[84], using Watson's transmission formula, has shown that Austin's results may be explained if it be assumed that the specific conductivity of the lower part of the ionosphere varies from sunspot maximum to sunspot minimum in the ratio 1·6 to 1. It will be seen that this figure is of the same order as that estimated by Chapman for the conductivity of the ionosphere as a whole.

Now it is well-known that there is a direct correlation between the incidence of magnetic storms and the sunspot cycle. It is also suspected that magnetic storms are due to the influx of material particles, originating from the sun, in the upper atmosphere. A very important question therefore arises. Is the variation in the ionospheric ionization due to an 11-year cycle in solar particle emission or due to a variation in the ultra-violet light which is now known to be mainly responsible for upper-atmospheric ionization?

Information bearing on this point can be obtained from measurements of upper-atmospheric ionization made by the critical penetration frequency method though, as such observations have been available only since 1931, a long series comparable to Austin's signal measurements is not yet available. Appleton and Naismith[85], making measurements of Region E ionization by the critical frequency method, found, from a comparison of the equinox values, a reduction in the ionization from 1931 to 1932 which they associated with the general fall in solar activity in the

sequence of the solar cycle. (The last sunspot maximum was in 1928 and the succeeding sunspot minimum in 1933–34.) They estimated the diminution from one year to the next to be of the order of 10 to 15 per cent and were inclined to interpret their results as indicating "quite a substantial variation in the sun's ultra-violet light during the solar period".

In view of the great reliability of Region F_1 critical frequency and also in view of the fact that this region, like Region E, is known to be caused by ultra-violet light, it would appear that a series of measurements of this type made over a duration of years would give the most reliable answer to the question posed above, but as yet no such series of values has been published.

The evidence derived from critical frequency measurements of Region F_2 at vertical incidence cannot be regarded as reliable until 1933 when the distinction between the ordinary ray and extra-ordinary ray critical frequencies was realized. Moreover, there is still the lack of agreement among different workers as to whether such critical frequencies really indicate an electron-limitation or an absorption-limitation effect. Gilliland[55], working at night when it is generally accepted that critical frequency measurements do indicate true values of ionization, found that the extra-ordinary ray critical frequency was greater in 1934 than in 1933, the provisional sunspot numbers for the two periods studied being 106 and 57 respectively. Eckersley[48] also made night-time measurements of critical frequency in transatlantic transmission of short waves and found the values of ionization greater on October 21–22, 1928, than on May 10–11, 1931. In view of the October-November day-time maximum of Region F_2 critical frequency, however, it does not seem that we can accept this as an unambiguous indication of a connexion with the sunspot activity.

Taking all the radio evidence together, it therefore appears probable that the ionization in Regions E and F_2 varies with solar activity, by way of a variation of the sun's ultra-violet radiation, but that a further series of measurements made by the critical frequency method, for Region F_1 especially, are required to set the matter beyond doubt.

The correlation of long-wave wireless transmission with solar activity established by Austin prompted other workers to examine the effects of a particular magnetic storm on radio phenomena. Observations on long waves (5000 metres) were first made by engineers of the Western Electric Company[86] who found that the effects of magnetic storms were to decrease greatly the night-time received signal and to increase slightly the day-time value. The corresponding question for short waves was studied by Appleton[87] who found that magnetic storms were associated with abnormally low values of signal intensity received from American short-wave stations in England. The latter result was confirmed by Eckersley[88] who also shows that transmissions to Canada were much more liable to be affected than transmissions to South Africa which agreed with the established result that magnetic storms are of greater magnitude in high than in low latitudes.

But signal intensity measurements cannot be regarded as unambiguous indices of ionospheric ionization, for it is well known that an increase of ionization may,

owing to absorption-limitation, result in a reduction of signal intensity, whereas in other cases such an increase, by the removal of electron-limitation, may cause just the reverse result. No such objection applies to measurements of ionization by means of the critical frequency method so that we may expect that the most reliable information should derive from its prosecution. Graphs of noon critical frequencies for England and America are shown in figures 8 and 9. It will be seen that in both cases there are day-to-day fluctuations. Schafer and Goodall[54] have found from their observations in America (see figure 9) that the value of Region F_1 ordinary ray critical frequency goes inversely with the magnetic activity as measured by the earth potentials at Deal. They find, however, no such correlation between magnetic activity and the critical frequencies for Region F_2.

The observations in England made by Appleton and Naismith[53] show that the critical frequencies of Regions F_1 and F_2 are subject to somewhat similar variations, and a comparison of figure 8 with a graph of magnetic activity shows that, in general, both $f_{F_1}{}^0$ and $f_{F_2}{}^0$ are low when the magnetic activity is high, though there are occasions when f_{F_2} is raised when magnetic activity is high.* The difference between the American and English results is not yet clear.

§ 11. ABNORMAL REGION *E* IONIZATION

It has been noticed by a number of observers that echo-reflections often occur from an atmospheric level approximating to that of Region E for electric-wave frequencies which are higher than the critical frequency for the normal Region E. A typical example of this effect is the occasional occurrence of echoes from an equivalent height of about 100 km. during the night at a time when the frequency used normally penetrates Region E. The phenomenon was reported by Appleton[89] in 1930, from a study of (P', t) records, who concluded that on such occasions "either the recombination of ions is prevented or there is some ionizing agent present which can influence the dark side of the earth". Similar occurrence of abnormal nocturnal Region E reflections was noted by Schafer and Goodall[90]. In 1932 Appleton and Naismith[85], using the critical frequency method, found that in south-east England an abnormally high value of critical frequency is often associated with magnetic activity. In the same year Ranzi[91], from a series of (P', t) observations, concluded that the abnormal night ionization in Region E is associated with the presence of barometric depressions either at the point of observation or at a place north of it.

Abnormal Region E reflection is, however, experienced in the day as well as during the night, and particularly is this the case in summer[85]. This result has been considered of interest in connexion with the suggestion of C. T. R. Wilson[92] that thunderstorms may contribute to the ionization in the upper-atmospheric conducting layer, in the form of either intense ionization currents or lightning flashes. In testing this point it has been found[85] that there is a relatively high correlation between the incidence of abnormal Region E conditions and a thunderstorm index

* This matter has been further discussed by Appleton and Ingram (*Nature*, **136**, 548 (1935)), who have attributed the abnormally low ionisation densities to the heating effect of the agency causing magnetic disturbances.

figure derived from observations on atmospherics. It has, however, since been found[53] that abnormal Region *E* ionization usually provides favourable transmission conditions for radio waves so that it is possible that the correlation found is to some extent an expression of the fact that atmospheric disturbances are relatively stronger under these abnormal conditions. The matter therefore cannot be considered as decided.

A further contribution to this subject has been made by Ratcliffe and White[93] who, from a rigorous statistical examination of (P', t) data, concluded that there was a definite connexion between magnetic storms and abnormal Region *E* ionization and also between such ionization and recorded thunderstorms in the British Isles. In connexion with these results and those of Ranzi, mentioned above, it may not be superfluous to recall the well-known connexion between thunderstorms and atmospheric depressions.

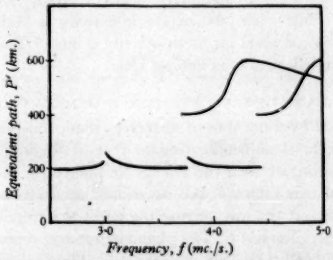


Figure 11. (P', f) relation for abnormal region *E* conditions.

Kirby and Judson[57] have made a study of abnormal Region *E* reflections in America and find no correlation between the occurrence of such ionization either with magnetic storms or with thunderstorms, the latter occurring either locally or within a radius of 300 km.

Skellett[94] has studied the relation between abnormal Region *E* ionization and the incidence of meteors and has observed (1) sudden increases in ionization in Region *E* lasting for a period of minutes or less, and (2) increases of longer duration, with maxima coincident in time with those of observed meteoric activity. Mitra Syam and Ghose[95] have also found the ionization in Region *E* abnormally high on the days of the Leonid meteoric shower of November 1933. The suggestion that meteors might influence radio wave transmission was first made by Nagaoka[96] in 1929.

Some idea of the nature and structure of the summer abnormal Region *E* has been obtained by the study of the relevant (P', f) records. Appleton and Naismith[53] have described a typical example, which is shown in figure 11. Here it is seen that as the frequency is increased from 2·6 Mc./sec. we get the normal type of (P', f) curve for Region *E*, the value of P' increasing rapidly as the penetration value

3·0 Mc./sec. for the ordinary ray is approached. For higher frequency values, however, we do not get the usual "jump" to Region F for there is a reflecting stratum, somewhere near the level of maximum ionization in Region E, which reflects waves of higher frequencies. In the example shown, waves of frequencies from 3·0 to 3·8 Mc./sec., ordinary ray, are reflected from it. The remarkable fall of P' for frequencies a little higher than 3·0 Mc./sec. may be explained as a group retardation effect in the normal Region E which is reduced as the frequency is increased. For a frequency of about 3·75 Mc./sec. it is seen that the "jump" to Region F occurs for the ordinary wave. This is, however, accompanied by a certain amount of overlap on the frequency scale indicating that the abnormal E stratum exhibits the phenomenon of partial reflection. It is also to be noted that such phenomena, observed for the ordinary ray, are similarly to be noted for the extra-ordinary ray at a range of frequencies 0·7 Mc./sec. higher. This may be interpreted as indicating that both the normal Region E and the abnormal stratum contain substantial electronic ionization.

§ 12. MEASUREMENT OF ELECTRON COLLISIONAL FREQUENCY IN THE IONOSPHERE

Eckersley[97] and Farmer and Ratcliffe[98] have made measurements of the collisional frequency of electrons in Region F. From (18) we have for temporal changes of the quantities

$$-\frac{\partial}{\partial t}(\log \rho)=\frac{\nu}{2c}\frac{\partial}{\partial t}(P'-P) \qquad \text{.....(42),}$$

where the differentiation refers to time t. By measuring corresponding changes of $\log \rho$ and P' with time Eckersley has, neglecting changes of P, estimated the value of ν in Region F_1 to be $3{\cdot}6 \times 10^3$ collisions per electron per second for an estimated height of 265 km. while Farmer and Ratcliffe, using a similar method, have found for Region F_2 a value of $1{\cdot}6 \times 10^3$ collisions per electron per sec.

§ 13. MEASUREMENT OF THE EARTH'S MAGNETIC FIELD IN THE IONOSPHERE

Appleton[21] has measured the earth's magnetic field in the ionosphere by measuring the night-time ordinary and extra-ordinary penetration frequencies for Region F. From (4) and (5) we have

$$H=\frac{2\pi m}{e}\left(\frac{f_E^2-f_0^2}{f_E}\right) \qquad \text{.....(43),}$$

from which H can be estimated. For south-east England the average value of H is found to be 0·42 gauss which is about 10 per cent less than the ground value (0·467 gauss).

With the observations made by Berkner and Wells[66] in Huancayo it is easy to show, using (43), that they indicate a value for H of 0·275 gauss which may be compared with the value of 0·296 at ground level.

§ 14. THE DIRECT-CURRENT CONDUCTIVITY OF THE IONOSPHERE

The first suggestion that the upper atmosphere was a pronounced electrical conductor was made by Balfour Stewart[99] from terrestrial magnetism considerations. His ideas in this connexion have been further developed by Schuster[100] and, in particular, by Chapman[101]. Estimates of the total conductivity (i.e. $\int \sigma \,.\, dh$ where σ is the specific conductivity, for horizontal electric forces, along a vertical element of height h) have been made by Schuster and Chapman. These indicate that the total conductivity must be of the order of 9×10^{15} e.s.u. Assuming, with Schuster, a total vertical thickness of 300 km., the specific conductivity must be of the order of 3×10^{8} e.s.u. which is of the same order as the specific conductivity of the ground.

The horizontal specific conductivity in our latitudes is largely determined by the earth's magnetic field. We may assume this, for simplicity, to be vertical. Now Townsend[102] has shown that the transverse direct current conductivity σ_τ is given, for such a case, by

$$\sigma_\tau = \frac{Ne^2}{m} \cdot \frac{\nu}{\nu^2 + p_H^2} \qquad \ldots\ldots(44),$$

where the medium has N charges, of mass m and charge e, per cm^3, ν is the collisional frequency and p_H is the angular gyro-frequency of the charges round the lines of magnetic force. In a uniformly charged atmosphere σ_τ will be a maximum when $\nu = p_H$ and at that level we have

$$(\sigma_\tau)_{max} = N \,.\, \frac{ec}{2H_0} = 14{\cdot}4\, N \qquad \ldots\ldots(45),$$

taking the earth's field as 0·5 gauss. It will thus be seen that the maximum specific conductivity is independent of the mass of the ions.

If the ionosphere contains both molecular ions and electrons there will be a maximum of electron-conductivity at a level where $\nu = p_H = eH_0/m_e c$, where m_e is the mass of an electron. The magnitude of this collisional frequency is thus $1{\cdot}32 \times 10^{6}$ c./sec. so that the conductivity maximum will probably occur at about the 90–100 km. level. Higher up, where $\nu = p_H = eH_0/m_i c$ (where m_i is the mass of an ion) there will be another maximum of conductivity, possibly at 200 to 300 km. in height. Since the two banks of conductivity, electronic and ionic, overlap, it seems fair to assume that the specific conductivity through the whole ionosphere is of the order of $14{\cdot}4\, N$ where N is the number of ions or electrons. Now in England N varies from $1{\cdot}1 \times 10^{5}$ electrons per cm^3 for Region E to $4{\cdot}5 \times 10^{5}$ electrons per cm^3 for Region F so that the conductivity would be from $1{\cdot}6 \times 10^{6}$ to $6{\cdot}5 \times 10^{6}$ e.s.u. It thus appears that if we consider the ionosphere to consist of equal numbers of ions and electrons the estimated electrical conductivity deduced from wireless observations is substantially less than the estimates made from the data of terrestrial magnetism. It should however be pointed out that, in the radio experiments, the quantity $\left(\frac{N_e}{m_e} + \frac{N_i}{m_i}\right)$ is measured so that the number of ions could be

at least 100 times the number of electrons without their being detected. An increase in the number of ions would bring about an almost proportionate increase in the estimate of transverse ionospheric direct-current conductivity. Whether the discrepancy is to be remedied in this way is not yet certain.

REFERENCES

Books

BOUTHILLON, L. *La Propagation des Ondes Électromagnétiques* (1921). Paris.
PEDERSEN, P. O. *The Propagation of Radio Waves* (1927). Copenhagen.
LABAT, P. *La Propagation des Ondes Électromagnétiques* (1932). Paris.
WHITE, F. W. G. *Electromagnetic Waves* (1934). London.

Papers

(1) ECCLES. *Proc. roy. Soc.* A, **87**, 79 (1912).
(2) LARMOR. *Phil. Mag.* **48**, 1025 (1924).
(3) APPLETON. *Proc. phys. Soc.* **37**, 22D (1925).
(4) NICHOLS and SCHELLENG. *Bell syst. Tech. J.* **4**, 215 (1925).
(5) BREIT. *Proc. Inst. Radio Engrs*, N.Y., **15**, 709 (1927).
(6) APPLETON. *Proc. Union Radio Sci. Int.*, Washington (Oct. 1927).
(7) GOLDSTEIN. *Proc. roy. Soc.* A, **121**, 260 (1928).
(8) HARTREE. *Proc. Camb. phil. Soc.* **27**, 143 (1931).
(9) BAKER and GREEN. *Rep. Aust. Radio Res. Bd*, No. 3 (1932).
(10) APPLETON. *J. Instn elect. Engrs*, **71**, 642 (1932).
(11) TAYLOR. *Proc. phys. Soc.* **45**, 245 (1933); **46**, 408 (1934).
(12) BOOKER. *Proc. roy. Soc.* A, **147**, 352 (1934); **150**, 267 (1935).
FÖRSTERLING and LASSEN, *Ann. Phys.*, Lpz., **18**, 26 (1933).
BAILEY. *Phil. Mag.* Ser. 7, **18**, 516 (1934).
(13) APPLETON and BUILDER. *Proc. phys. Soc.* **45**, 208 (1933).
(14) APPLETON. *Nature*, Lond. (Feb. 7, 1931).
See also APPLETON and NAISMITH, *Proc. roy. Soc.* A, **137**, 36 (1932).
(15) —— *Proc. phys. Soc.* **41**, 43 (1928); **42**, 321 (1930).
(16) DE GROOT. *Phil. Mag.* **10**, 521 (1930).
(17) APPLETON. *Nature*, Lond., **135**, 618 (April 20, 1935).
(18) HARTREE. *Proc. Camb. phil. Soc.* **25**, 97 (1929); **27**, 143 (1931); and *Proc. roy. Soc.* A, **131**, 428 (1931).
(19) APPLETON and BARNETT. *Nature*, Lond. (March 7, 1925); and *Proc. roy. Soc.* A, **109**, 621 (1925).
(20) BREIT and TUVE. *Phys. Rev.* **28**, 571 (1926).
(21) APPLETON. *Nature*, Lond., **133**, 793 (May 26, 1934).
(22) GILLILAND and KENRICK. *Bur. Stand. J. Res.*, Wash., p. 783 (Nov. 1931).
WHITE, E. L. C. *Nature*, Lond., **129**, 579 (1932).
RUKOP and WOLFF. *Z. tech. Phys.* **132** (March 1932).
GOUBAU and ZENNECK. *Hochfrequenztech. u. Elektroakust.* **41**, 77 (1933).
(23) GILLILAND. *Bur. Stand. J. Res.*, Wash., **11**, 501 (Oct. 1933).
NAISMITH. *Nature*, Lond., **133**, 66 (1933).
(24) BREIT. *Proc. Inst. Radio Engrs*, N.Y., **17**, 1815 (1929).
(25) APPLETON. *Proc. roy. Soc.* A, **128**, 548 (1930).
(26) APPLETON and BARNETT. *Proc. roy. Soc.* A, **109**, 621 (1925).
SMITH-ROSE and BARFIELD. *Proc. roy. Soc.* A, **110**, 580 (1926).
(27) MIRICK and HENTSCHEL. *Proc. Inst. Radio Engrs*, N.Y., **17**, 1034 (1929).
(28) SMITH-ROSE and BARFIELD. *J. Instn elect. Engrs*, **64**, 831 (1926).
(29) RATCLIFFE and PAWSEY. *Proc. Camb. phil. Soc.* **29**, 301 (1933).
PAWSEY. *Proc. Camb. phil. Soc.* **31**, 125 (1935).

(30) MARTYN and GREEN. *Proc. roy. Soc.* A, **148**, 104 (1935).
(31) PARKINSON. *Proc. Inst. Radio Engrs*, N.Y., **17**, 1042 (1929).
(32) CHERRY and MARTYN. *Rep. Austr. Radio Res. Bd*, No. 4 (1932).
(33) APPLETON and RATCLIFFE. *Proc. roy. Soc.* A, **117**, 576 (1928).
(34) APPLETON and WATT. *Wireless World*, **17** (July 8, 1932); *Proc. Phys. Soc.* **45**, 211 (1933).
(35) RATCLIFFE and WHITE, E. L. C. *Phil. Mag.* Ser. 7, **16**, 125 (1933).
(36) ECKERSLEY. *Nature*, Lond., **129**, 579 (1932).
(37) RATCLIFFE and WHITE, F. W. G. *Phil. Mag.* Ser. 7, **16**, 423 (1933).
(38) APPLETON. *J. Instn elect. Engrs*, **71**, 430, 645 (1932).
(39) HOLLMANN and KREIELSHEIMER. *Elekt. Nachr.-Tech.* **10**, 392 (1933).
(40) WHITE, F. W. G. *Proc. phys. Soc.* **46**, 91 and 805 (1934).
(41) APPLETON. *Nature*, Lond., **120**, 330 (1927); *Proc. roy. Soc.* A, **128**, 159 (1930).
APPLETON and GREEN. *Proc. roy. Soc.* A, **128**, 160 (1930).
(42) SCHAFER and GOODALL. *Proc. Inst. Radio Engrs*, N.Y., **19**, 1434 (1931).
(43) APPLETON and BUILDER. *Proc. phys. Soc.* **44**, 76 (1932).
(44) ECKERSLEY. *J. Instn elect. Engrs*, **71**, 423 (1932).
SCHAFER and GOODALL. *Nature*, Lond. (June 3, 1933).
APPLETON. *Nature*, Lond. (June 17, 1933); *Proc. phys. Soc.* **45**, 673 (Sept. 1933).
KIRBY, BERKNER, GILLILAND and NORTON. *Bur. Stand. J. Res.*, Wash., **11**, 829 (1933).
(45) APPLETON and BUILDER. *Proc. phys. Soc.* **44**, 81 (1932).
(46) MIMNO. *Nature*, Lond., **134**, 63 (1934).
(47) KIRBY and JUDSON. *Bur. Stand. J. Res.*, Wash., **14**, 470 (1935).
(48) ECKERSLEY. *J. Instn elect. Engrs*, **71**, 429 (1932).
(49) CHAPMAN. *Proc. phys. Soc.* **43**, 26 (1931); **43**, 483 (1931).
(50) MILLINGTON. *Proc. phys. Soc.* **44**, 580 (1932); **47**, 263 (1934).
(51) APPLETON and NAISMITH. *Proc. roy. Soc.* A, **137**, 36 (1932).
(52) BUILDER. *J. Instn elect. Engrs*, **71**, 456 (1932).
(53) APPLETON and NAISMITH. *Proc. roy. Soc.* A, **150**, 685 (1935).
(54) SCHAFER and GOODALL. *Proc. Inst. Radio Engrs*, N.Y., **23**, 674 (1935).
(55) GILLILAND. *Bur. Stand. J. Res.*, Wash., **11** (Oct. 1933); **14**, 283 (1935).
(56) KIRBY, BERKNER and STUART. *Proc. Inst. Radio Engrs*, N.Y., **22**, 481 (1934).
(57) KIRBY and JUDSON. *Bur. Stand. J. Res.*, Wash., **14** (April 1935).
(58) APPLETON. *Phys. Rev.* **47**, 89 (1935); **47**, 704 (1935); *Nature*, Lond., **136**, 52 (1935).
(59) HULBURT. *Phys. Rev.* **31**, 1018 (1928); **35**, 240 (1930); **39**, 977 (1932).
(60) BURROWS. *Proc. Inst. Radio Engrs*, N.Y., **19**, 1634 (1931).
(61) APPLETON, NAISMITH and BUILDER. *Nature*, Lond., **132**, 340 (1933).
(62) WAGNER. *S.B. Preuss. Akad. Wiss.* **32**, 1 (1933).
(63) FLEMING. *Terr. Magn. atmos. Elect.* p. 305 (Dec. 1934).
(64) BONTCH-BRUEWITCH. *Proc. Inst. Radio Engrs*, N.Y., **22**, 1124 (1934).
(65) RANZI. *Nature*, Lond., **133**, 29 (Jan. 1934).
(66) BERKNER and WELLS. *Terr. Magn. atmos. Elect.* p. 215 (Sept. 1934).
(67) HULBURT. *Phys. Rev.* **46**, 822 (1934).
(68) WELLS. *Terr. Magn. Atmos. Elect.*, p. 209 (Sept. 1934).
(69) RACKSHIT. *Phil. Mag.* **12**, 897 (1931).
(70) MITRA and RACKSHIT. *Phil. Mag.* **15**, 20 (1933).
(71) GREEN. *Rep. Aust. Radio Res. Bd*, No. 2, Melbourne (1932).
(72) WHITE, F. W. G. *Proc. phys. Soc.* **46**, 91 (1934).
(73) APPLETON. *Proc. phys. Soc.* **45**, 673 (1933).
(74) APPLETON and CHAPMAN. *Nature*, Lond. (May 21, 1932).
(75) CHAPMAN. *Mon. Not. R. astr. Soc.* **92**, 413 (1932).
(76) SCHAFER and GOODALL. *Nature*, Lond., p. 393 (March 9, 1935).
(77) For references to work done in connexion with eclipses (1) to (4) inclusive, see APPLETON and CHAPMAN, "Report on ionisation changes during a solar eclipse", *Proc. Inst. Radio Engrs*, N.Y., **23**, 658 (1935).

(78) ECKERSLEY. *J. Instn elect. Engrs*, **71**, 405 (1932).
(79) APPLETON. *Proc. phys. Soc.* **45**, 673 (1933).
(80) CHAPMAN. *Proc. roy. Soc.* A, **132**, 353 (1931).
(81) —— *Philos. Trans.* **218**, 113 (1917).
(82) AUSTIN. *Proc. Inst. Radio Engrs*, N.Y., **12**, 521 (1924).
(83) AUSTIN and WYMORE. *Proc. Inst. Radio Engrs*, N.Y., p. 166 (Feb. 1928).
(84) APPLETON. *Second Report on Solar and Terrestrial Relationships, International Research Council*, p. 18 (Paris, 1929).
(85) APPLETON and NAISMITH. *Proc. phys. Soc.* **45**, 389 (1933).
(86) See ANDERSON. *Proc. Inst. Radio Engrs*, N.Y., **16**, 297 (1928).
(87) APPLETON. *Electrician*, p. 256 (March 11, 1927).
(88) ECKERSLEY. *J. Instn elect. Engrs*, **67**, 1019 (1929).
(89) APPLETON. *Proc. roy. Soc.* A, **126**, 567 (1930).
(90) SCHAFER and GOODALL. *Proc. Inst. Radio Engrs*, N.Y., **20**, 1131 (1932).
(91) RANZI. *Nature*, Lond., **130**, 368 (1932).
(92) WILSON. *Proc. phys. Soc.* **37**, 32 D (1925).
(93) RATCLIFFE and WHITE. *Proc. phys. Soc.* **46**, 107 (1934).
(94) SKELLETT. *Proc. Inst. Radio Engrs*, N.Y., **23**, 132 (1935).
(95) MITRA SYAM and GHOSE. *Nature*, Lond., **133**, 533 (1935).
(96) NAGAOKA. *Proc. imp. Acad. Japan*, **5**, 233 (1929).
(97) ECKERSLEY. *Nature*, Lond., **135**, 435 (1935).
(98) FARMER and RATCLIFFE. *Nature*, Lond., **135**, 485 (1935).
(99) BALFOUR STEWART. *Encyc. Brit.* Art. "Terr. Mag." (9th edition, 1882).
(100) SCHUSTER. *Philos. Trans.* A, **208**, 181 (1907).
(101) CHAPMAN. *Philos. Trans.* A, **213**, 279 (1913); **214**, 295 (1914); **215**, 161 (1915); **218**, 1 (1919).
(102) TOWNSEND. *Electricity in Gases*, Oxford.

SOUND

By E. G. RICHARDSON, B.A., Ph.D., D.Sc.

§ 1. FREE PROPAGATION: SUPERSONICS

The bulk of research on the propagation of sound continues to lie in the region above the audible pitch limit. While the Pierce interferometer remains the favourite apparatus for gases, light diffraction by the supersonic radiation is more generally used for liquids. Discussion on the theory of the interferometer has considerably abated. In this connexion we have only to note that Grossmann[1] has considered the effect on the wave-length near the source of the spreading which takes place there. He has obtained a formula which allows one to correct the observed distance between peaks near the source for this effect. Out of courtesy, we will award the place of honour in our account of experimental research to our friends the physical chemists who have seized on this new technique with avidity, though not always with judgment. Measurements of wave-lengths in gases have been continued by Spakovski[2] (carbon dioxide), Richards and Reid[3] (carbon dioxide, carbon bisulphide and sulphur dioxide), Eucken and Becker[4] (carbon dioxide and chlorine in mixtures with other gases), Kneser and Wallmann[5] (hydrogen-ethylene mixtures) while Curtis[6] and Rogers[7] have traced the absorption in carbon dioxide, the latter in addition noting the effect of humidity on this absorption. Eucken and Becker have attempted an ambitious programme involving measurements of wave-length in two principal gases, with admixture of five more, at several concentrations, as well as the effect of temperature on dispersion in the principal gases, but unfortunately their range of oscillators is very limited. Three frequencies in fact seem to them sufficient to characterize the dispersion curves, and in a number of cases only one is used. The reader will probably now be familiar with the S-shaped curve which is characteristic of the trend of velocity with frequency in the dispersion region in such gases as carbon dioxide, but he will experience a shock on looking at their dispersion figures where the S-curve is sometimes drawn through a *single plotted point*. Life is not so simple in the realm of pure physics! However, the authors succeed in extracting from their material much information on the behaviour of the molecules of the gases in collision and their specific heats which should be of interest to the chemist. The work of Richards and Reid is important, inasmuch as they give the first measurements and interpretation of the effect of pressure and temperature on supersonic dispersion. They have indicated within a certain range of frequency (n) that the dispersion curves at different temperatures (T) and pressures (p) are reducible to a single curve if parameters n/p or nT are employed. Although they lack the temerity of the last

mentioned authors their work might also be criticized in respect of paucity of frequencies, so that it becomes difficult to judge how well the observations submit to this re-shuffling.

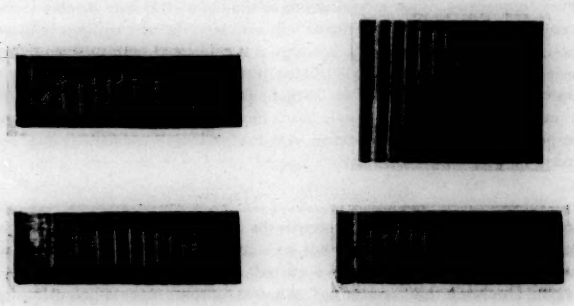
This work is followed up by a discussion in the new *Journal of Chemical Physics* on the theoretical interpretation of dispersion in terms of activation-energy of molecules, in which Rose[8], Richards[9], and Huntington[10] take part. The discussion leads them further and further from experiment when they have to invoke three or five of these levels, each involving a step-up in velocity too slight to be measurable with present supersonic technique. The relaxation-time theory (cf. last *Report*) requires that once these jumps are achieved the velocity should remain at the new level. The writer[11], using a method in which a hot wire is traversed in the space between source and reflector, showing maxima (at the nodes) and minima (at the antinodes), has shown that the wave-length in a dispersive gas diminishes at higher frequencies. What modifications to the theory this may involve is not yet apparent. The hot-wire measurements will also give the absorption coefficient of the gas. In these experiments the hot-wire was calibrated both in steady currents of the gas in question, and also by being traversed through the stationary wave system of a non-dispersive gas, viz., argon, in which the velocity from point to point was assumed to vary in a simple harmonic fashion. From the size and position of the peaks, the attenuation constant α and wave-length constant β may be calculated. The curves have the form

$$\cosh\{2\alpha(x-l)\}-\cos\{2\beta(x-l)\},$$

x being measured from the quartz towards the reflector and l being the total length of the interspace. Traverses of the hot-wire across the face of the quartz showed that more spreading of the radiation occurred in carbon dioxide, for example, than in oxygen. Doubtless this "scattering" accounts for some of the apparent absorption in such a gas. Sherratt and Griffiths[12] use the Pierce apparatus in carbon monoxide in a furnace at temperatures between 1000° and 1800° C. The gas is contained in a carbon tube, 3 ft. long and $\frac{1}{2}$ in. in internal diameter. This involved a tube correction to the observed velocity of sound, for which the usual Kirchhoff formula was adequate. It amounted to between $\frac{1}{2}$ and 1 per cent. Corrections were also made for expansion of the tube and the screw holding the reflector and for impurity of the gas, which could contain from 0·3 to 0·7 per cent of nitrogen. A correction to the specific heat was made for possible dispersion, in accordance with the Kneser theory, although only two frequencies (7·9 and 27 kc./sec.) were used. At ordinary frequencies Keesom and van Lammeren[13] have continued their researches on gases at low temperatures in their resonator apparatus, by determining the ratio of the specific heats of neon down to its critical temperature. Swanson[14] with the same object in view has used his interferometer (described in the previous *Report*) on a number of organic liquids at pressures between atmospheric and 300 kg./cm.^{2} The velocity increases in linear relationship to the pressure over this considerable range.

Apart from this last mentioned research, workers on supersonics in liquids have

continued to favour the optical diffraction method, originally suggested by Brillouin, and fully described in the last *Report*. In particular, Bachem, Hiedemann and their co-workers in Cologne have produced no less than ten papers in the year, in which the Brillouin method is the basis. The principal advance in technique which they have made, is to render a whole area of the field of supersonics in the liquid visible by optical diffraction of a beam of light which cuts across it at right angles. This device is of most avail when stationary waves are formed between the quartz and a reflector, plane or curved[15,16,17], but if a stroboscopic interruption of the light at the same frequency as the supersonics is arranged, a section of a field of *progressive* waves may be photographed[18,19]. The stroboscopic effect is obtained by passing a beam of polarized light through a Kerr cell actuated by alternating current derived from the oscillator maintaining the quartz. Other effects which may be conveniently studied with this new technique are (*a*) the abnormal velocity



[*Reproduced by kind permission of Prof. E. Hiedemann.*]

Figure 1. Progressive waves produced by supersonic radiation in liquids.

near the sender[20], (*b*) the oscillations of the quartz itself[21,22], (*c*) interference effects with supersonic waves[23]. Figure 1 is from actual photographs illustrating the first of these effects at a frequency of 3284 kc./sec. We are looking at the tray of liquids between the quartz (on the left) and a reflector (out of view on the right), and the stationary waves in the liquid have at each nodal line diffracted the beam of light coming from behind to form dark lines. Thus the distance between consecutive dark lines gives the half wave-length of the supersonics in the liquid. The abnormal wave-length near the sender is quite obvious. This has been noticed by more workers on the subject but not demonstrated visually before. The authors here ascribe it to heating effect of the radiation producing a temperature gradient in the liquid near the source. It is only fair to add that the Kerr-cell stroboscope and its application to the study of *progressive* waves was suggested by Debye at a Berlin conference in 1933, and has also been realized by Sach and Coulon[24] under his guidance. Lucas[25] and Baumgardt[26] have suggested other modifications of the Brillouin method which have not yet been realized in practice.

Bär[27] and Walti[28] show that if the source has adequate intensity it is possible to transfer sufficient of the energy in an irradiated liquid to a solid body immersed in it, so that the velocity of transverse or longitudinal waves in the latter may be measured. Actually the coefficient of transmission of the supersonics through the solid in the form of a wedge is measured. The transmission is a maximum for thicknesses which are a multiple of the half-wave-length. The two velocities, for transverse and longitudinal waves having thus, in combination with the Brillouin effect in the liquid, been measured, Poisson's ratio for the plate may be calculated. Sack[30] uses a similar device for the absorption in solid specimens such as lead, steel and salt crystals. Wyss[29] uses the supersonic interferometer in liquids but observes when the light-diffracted intensities pass through maxima as the screw is moved, to get the wave-length, thus neatly combining the two chief supersonic instruments.

Passing now to the miscellaneous effects of supersonics we find an exhaustive study by Claus[31] on their effect on photographic emulsions. He finds that after irradiation for a considerable time the stability and homogeneity of the emulsion is improved. This reminds us of the attempt which has been reported in the press to improve the quality and stability of another emulsion, viz. milk. In the photographic emulsion the silver-halogen concentration can be increased above the normal after irradiation, and a general tonic effect on the sensitizing process results. It seems as though this work will have important consequences for the photographic industry. In connexion with the splitting up of starch and other complex molecules, Szalay[32] has worked out the energy of oscillation which is available in the supersonic radiation. In an actual experiment he estimated, from the rate of heating of water, the amplitude of vibration to be 10^{-4} cm., while the acceleration came out at 2.10^9 cm./sec.^{2} If starch molecules in solution take up this same amplitude, they must be endowed with a very high energy, on account of their comparatively large mass; indeed if all the energy turned to heat it would cause a rise of temperature of over 100° C. Such a large amount of energy in the molecule might be sufficient to break it down. In the same direction, Liu and Wu[33] state that the oxidation of certain substances in solution which has been observed under the action of supersonics is necessarily accompanied by cavitation and that it must be the oxygen in the bubbles so forced out of solution that is activated and so inspires reactions such as the liberation of iodine from iodide, which has been previously observed. We must leave to the chemists the discussion of the feasibility of this reaction. We must also mention that it has been possible to destroy a tobacco virus held in suspension in oil, by this means, but it seems scarcely feasible to make economic use of this process.

Both Rocard[34] and King[35] have examined the radiation field from an oscillator such as the piezo-electric quartz from the point of view of the effect of viscosity in damping the radiation. While the former shows that even at sonic frequencies this viscous damping may be quite comparable with any anomalous absorption which may be observed, the latter is concerned with the distance to which a supersonic beam may penetrate in water as in the well-known echo-sounding apparatus. For

an oscillator of diameter a working at a frequency $n=kc/2\pi$, the ratio of displacement amplitude at a distance r to the velocity amplitude of the source is given by

$$\tfrac{1}{2}a^2ke^{-\kappa r}/r,$$

where κ is the appropriate coefficient of damping ($=\frac{2}{3}\gamma k^2c$, γ being the dynamical viscosity). This expression has a maximum value at a given value of r for a wavelength

$$\lambda=4\pi\,(\gamma r/3c)^{\frac{1}{2}}.$$

With $r=4\times10^5$, $\mu=0{\cdot}01$, King finds $\lambda=1{\cdot}48$ cm., corresponding to a frequency of 100 kc./sec. To every distance there is in fact an optimum frequency of transmission, an experimental fact which was observed by Boyle in 1918. The intensity of sound radiation in different azimuths relative to a spark-gap source has been investigated by Skomtao and Su[36] and also by McFarlane[37]. The former used a Rayleigh disc, the latter a torsion vane, to measure the intensity. Both found that the intensity was not uniformly distributed but, as perhaps might be expected, was more conspicuous in an equatorial plane; they obtained in fact a polar curve very like that for the distribution of illumination round an arc lamp. Besides a theoretical paper due to Biot[38] on the acoustic spectrum of a body submitted to a shock such as a struck string, we must note two attempts to analyse the spectrum of such a source. Hickman[39] passes the electric current from a microphone, which picks up the sound, through an electromagnet which confronts a set of tuned reeds. These respond in such a way that the amplitude of oscillation of a light image cast on a screen by mirrors on the reeds is proportional to the amplitude of the corresponding component in the driving circuit. It is therefore an analyser of the Hartmann-Kempf type. Analyses of such sounds by this apparatus, which has a five octave range and 48 reeds to an octave are given in the paper. On the other hand, the acoustic spectrometer of Meyer and Thienhaus[40] follows the Dvorak type (and the optical spectrometer) in having a grating of steel nails 3·4 mm. in diameter arranged round an arc 3 metres long of a circle of the same radius, with such precision that each nail was assured to $\frac{1}{50}$ mm. Truly a test of patience! A duralumin ribbon sender was used as source, and the nail-system used as a reflection grating. A condenser microphone which could be swung on the end of a lever through the series of foci like the telescope in the optical instrument served as detector. Specimen spectra are given by the authors and reproduced in figure 2.

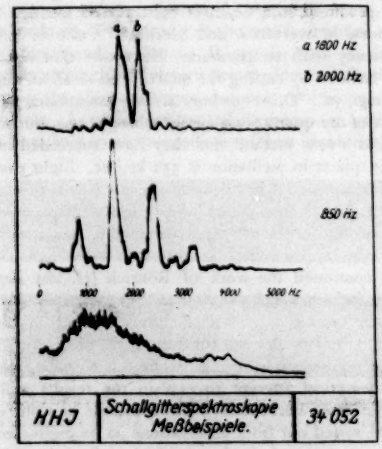
Besides a theoretical paper by Young[41] on the Doppler effect in a moving medium, there is a remarkable theoretical paper on acoustic radiation pressure in which King[42] discusses the formation of striations by dense particles in a fluid under supersonic radiation. His results for the pressure experienced by a small sphere in a plane progressive or stationary wave should be of considerable importance in practice. Thus a small sphere of radius r and density ρ_1 in a progressive wave of wave-length constant $k\,(=2\pi/\lambda)$, and amplitude a travelling through a medium of density ρ_0 experiences a mean pressure

$$\overline{P}=2\pi\rho_0\,|\,a\,|^2\,k^6r^6\,\{1+\tfrac{2}{9}\,(1-\rho_0/\rho_1)^2\}/(2+\rho_0/\rho_1)^2.$$

The "smallness" of the sphere is, as often in such problems, limited by $kr=1$.

"It is interesting to note that the relative density factor (in brackets) is always finite and positive and has a minimum at $\rho_0/\rho_1 = 3$."

If we are endeavouring to detect supersonics in water using a sphere of equal density in the form of a torsion balance ($\rho_0/\rho_1 = 1$) at optimum sensitivity, it is calculated that the intensity level must exceed 100 decibels for success. In a plane stationary wave, $\phi = a \cos k(x+h)$ the radiation pressure is periodic, depending on



[*Reproduced by kind permission of Dr Erwin Meyer.*]

Figure 2. Sound spectrographs (1) two pure tones, (2) a complex note, (3) a noise.

the position, h, of the centre of the sphere with respect to the nodal planes. For small spheres

$$\overline{P} = \pi\rho_0 \mid a \mid^2 \sin 2kh \,.\, k^3 r^3 \{1 + \tfrac{2}{3}(1 - \rho_0/\rho_1)\}/(2 + \rho_0/\rho_1).$$

This is larger, *ceteris paribus*, than in a progressive wave; further if $\rho_0/\rho_1 < 2{\cdot}5$, the relative density factor is positive but if $\rho_0/\rho_1 > 2{\cdot}5$ it is negative. This illustrates the well-known fact that in Kundt's tube the dense particles are urged away from the nodes, the light particles on the contrary towards them, the stable position in the former case being at the antinodes (cf. p. 78 of last *Report*). The particles tend to take up the oscillation to an extent dependent upon the relative density—the function in this case being $(2 + \rho_0/\rho_1)/\{(\rho_0/\rho_1)(5 - 2\rho_0/\rho_1)\}^{\frac{1}{2}}$—and their initial distances from an antinode. King's deductions on this score are confirmed by earlier observations of Boyle on particles dropped into a supersonic stationary wave beam.

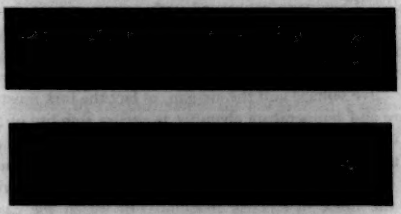
§ 2. VIBRATING SYSTEMS: SOLIDS

We find ourselves still in the supersonic field when we examine the work on solid bodies in vibration. Osterburg[43] continues his studies on the interference patterns in light reflected from quartz oscillators in vibration, using an improved optical interferometer of the Michelson type, in which the quartz causes periodic changes in the optical path length. He is joined by Bergmann[44] who uses the diffraction bands produced in a beam of light passed through the quartz as an indicator of its possible overtones, and Straubel[45] who is concerned with the variation of frequency with temperature. He shows that a small temperature coefficient can be ensured by cutting the quartz at as small an angle as possible to the optical axis, e.g. 30°. Other authors already mentioned have deduced the modes of vibration of the quartz from similar observations, but with Petrzilka and Zachoval[46] we find a new method, for they have succeeded in taking *schlieren* photographs of the quartz in oscillation at 552 kc./sec. Light passing through the quartz and falling on a screen is just cut off, where it passes through a focus, by an obstacle pushed in from one side. When the quartz is set in piezo-electric oscillation an image appears on the screen which, owing to the change in double refraction suffered by the light, appears dark at the nodes and light at the antinodes. Schoeneck[47] has continued the work of Röhrich (cf. last *Report*) on rods in vibration under the influence of loose contact with quartz oscillators, using single crystals of zinc, cadmium and tin. They all show dispersion in the same sense that the velocity of the various modes is a function of the frequency, but it was found so difficult to classify these modes and to excite one without interaction from the others that any theoretical attempt to explain the results will be exceedingly difficult. In simpler cases Bechmann[48] has indicated how the elastic properties of the body may be sorted out from observations on its piezo-electric frequencies. The remaining papers on rods in longitudinal and torsional vibration are by Cremer[49] and Schenk[50] respectively, and are mainly of theoretical interest. The former has examined the resonance curves of a rod suspended at the end by threads and excited by a periodic force at one end, while the latter is concerned with the possible change of rigidity at high frequencies of excitation. Both show that dispersion may occur in the two types of oscillation they have examined.

In the technique of performing the Fourier analysis of the transverse vibrations of a solid body, we note an improvement in the stroboscope due to Nemes[51]. Instead of causing the intermittent illumination of the object to take place discontinuously in the on-off-on-off manner of the usual crenellated disc he produces a sinusoidal illumination by having the light projected through a template cut to a sine wave. This template forms the circumference of the disc which is rotated. If the beam of light now falls on a conventional patterned disc, rotated at the same speed, that pattern which is repeated as many times round the circumference as the template has sine waves will appear stationary. If the template contains in addition the nth harmonic, the pattern which is repeated n times as often will also appear stationary. That is how the apparatus is arranged for the demonstration of Fourier

analysis. In application, the vibrating reed or other solid would take the place of the patterned disc. Apparently the disadvantage of the apparatus lies in the number of templates which would need to be provided, but the pure sine-wave form of the luminous flux may also be modulated by cutting off part of the source so that only the peaks are illuminated, giving the Fourier series of asymmetrically distorted sine waves.

Passing on to strings, we find two remarkable investigations of the acoustics of two musical instruments, viz. the violin and the pianoforte. Lark-Horovitz and Caldwell[52] have made an X-ray study of the wood structure of famous violins including a Stradivarius and an Amati. It appears that the front of the belly (usually of spruce) on which the strings are stretched has a definite fibrous structure in a good violin and is therefore anistropic to sound-wave propagation, but the back is best free of structure. As far as they can tell at present the quality of the varnish, which has always been held by connoisseurs to play an important part in producing



[*Reproduced by kind permission of Mr Hart.*]

Figure 3. Oscillograms of piano strings—the upper one of each pair as produced by the artist, the lower as produced by the pendulum.

evenness of quality, is unimportant. The further work which they promise on this question will be awaited with interest. Two decades ago the Physical Society held a discussion on that elusive factor in pianoforte technique called "touch". The opinion prevailing then among the disputants was that it was a fetish set up by the virtuosi for their public to worship. This view has received conclusive confirmation from the work of Hart, Fuller and Lusby[53] at the University of Pennsylvania. The question they asked themselves was this: Is there any peculiar quality of tone which, let us say, a Paderewski or a Schnabel can produce from a given piano which cannot be produced from the same piano by "the little girl next door" at her first or second lesson, or even by a mechanical player? To answer this question they made oscillograph records of the notes produced by two American virtuosi, and a set of records of the notes produced by a pendulum which on release, hit the same piano key. In every case it was possible to get an exact match of one of the human records with one of the robot's. Two pairs of these tallies are shown in figure 3. The inference is, in the words of a speaker at the 1913 discussion, "The only factor that a player can vary is the speed with which the hammer hits the

string". Ghosh[54,55] also has considered the struck string from the theoretical standpoint of a collision problem between the soft felt hammer and the hard string. He has evolved a generalized theory of which, he claims, the earlier theories of Helmholtz, Kaufmann, etc. were special cases.

In a noteworthy publication, Davies and James[56] report an exhaustive theoretical and experimental investigation of the electrically maintained reed. They are concerned with the accuracy with which Young's modulus can be found from a delineation of the resonance curve of the reed. Their theory embraces such refinements as finite amplitude, hysteresis, and eddy currents in the core. For the experiments they chose a phosphor-bronze reed with stalloy armature excited by a valve oscillator, measuring the motional impedance of the system at various frequencies on an A.C. bridge. Finally they are able to quote Young's modulus for phosphor bronze to 1 in 1000, which compares favourably with the accuracy of the statical (deflection) method.

An equally painstaking research on the *valve maintained* tuning fork is due to Dye and Essen[57]. They were concerned with constructing a frequency standard and succeeded in establishing the frequency to one in a million. Balancing the prongs is one important consideration. Unless this is carefully done, the frequency will change when the base is clamped to a bench. Other factors which must be held constant are the valve voltage and the air-gap, in fact the fork must be enclosed in a constant pressure and temperature chamber to secure stability of frequency.

Also in this domain are the papers of Araki[58] who has worked out the intensity distribution in the neighbourhood of a vibrating tuning fork; and Brown[60] who has considered the incomplete ring in transverse vibration, with the view to applications to the noise of electric machines, in which the case enclosing the stator may be idealized to a ring-shaped strip with its ends nearly touching. While Ruedy[61] has considered the radiation from a membrane vibrating in an overtone and compared it with that of a rigid oscillating piston from the point of view of loud-speaker performance, Raman[62] has dealt with the properties of an Indian native drum called "midranga" which instead of having a skin of uniform thickness has one painted with a paste of charcoal and starch in such a fashion that the "mass per unit area" progressively increases from the circumference to the centre. The result is a rich and harmonic series of overtones, of which Raman has obtained the nodal patterns.

Colwell and Bryant[63] have published some very clear photographs of Chladni figures on plates vibrating in high overtones, the plates being excited by contact with nickel rods set in magneto-strictive oscillation.

The bell continues to invite research work by its very complexity. Obata and Tesima[64] have investigated the acoustics of Japanese hanging bells, whose tone quality differs materially from that of Western bells. They consist of a deep cylindrical shell ending in a hemispherical cap and are more uniform in thickness. While the inside is smooth, the outside is heavily embossed and includes striking points in the form of the sacred Buddhist symbol—a lotus flower. The external striker is a heavy piece of wood. The lower components have frequencies in the

ratios 2 : 5 : 7 : 9 : 11, but many of them are not pure tones. For instance the "fundamental" consisted of two tones 246 and 248 producing a very obvious slow beat. The authors attempt to correlate the frequencies with the possible positions of nodal line of latitude and longitude. Attempts to imitate bell-tone by gongs made of steel wires bent into spirals and struck at their free ends have been made by clockmakers with a certain amount of success. Investigations have been made by Pomp and Zapp[65] to determine what is the best type of metal for the purpose. Carbon steel (0·63 per cent) tempered for 15 minutes at 300° C. gave the maximum duration of note.

Another suggestive idea in the practice of carillon ringing is due to Curtiss and Giannini[66]. The main objection to large bells in a carillon are the cost of the bells themselves and the heavy tower to hang them in, together with non-uniformity of quality throughout the peal. The new suggestion is to use small bells, 3 to 15 in. high and to amplify them on a loud speaker of some 200 watts. Tests with a number of small bells so amplified to form a peal are made possible owing to the low initial cost of the bells.

In this section we may include an extension by Mary Waller[67] of her work on the Trevelyan rocker (cf. p. 74 of last *Report*). In confirmation of the suggestion that the sound heard when hot metals are placed in contact with solid carbon dioxide is largely assisted by the coercion of the gas evolved, a *schlieren* photograph of the action has been made which shows up the sublimation process very well. It is proposed that the "vibrating properties" of metals should be studied in this manner.

§ 3. VIBRATING SYSTEMS: FLUIDS

The revived interest in the sensitive flame which we remarked upon in the last report continues with unabated vigour. Carrière[68] has entered this field with stroboscopic photographs of such flames, which cannot be said to improve upon earlier ones made by other investigators. He has however made a variety of types of flame and has adapted the sensitive flame to test when a heterodyne wave-meter is in unison with the oncoming Hertzian waves.

Zickendraht[69] continues his earlier work by making measurements of the distribution of temperature and velocity in sensitive flames. His colleague Harding[70] is able to show that two sensitive flames have this in common with the pair of pulsating spheres studied by Bjerknes, that a mutual reaction occurs, the flames repelling each other if acted upon by periodic forces 90° out of phase with each other, but attracting each other when the disturbing forces are in phase. Zickendraht attempts a theory of the sensitive flame on this foundation but his attempt at a theory of the flame does not take us far on the road to the explanation of this intriguing phenomenon.

Meier[71] applies the sensitive flame to a familiar experiment, the focussing of sound from a whistle by parabolic mirrors; neither can a claim for novelty be put forward for his demonstration that the sound of a whistle reflected back upon itself causes a reduction in its intensity. This is the principle of the supersonic

interferometer over again. Of more interest is the observation of Hopfield[72] that when two high-speed water jets collide at an angle a note is given out whose frequency as shown by stroboscopic illumination is determined by the rate of formation of droplets in the spray. When the water is replaced by gas a supersonic vibration takes place.

We have a considerable amount of research on tubes and columns of air to report in 1934. On the theoretical side Press[73] has dealt with the case in which sound is radiated through the walls of a tube and has shown that both stationary and progressive waves may be expected in the column and in the walls due to the "leakage" of the latter. Lehmann[74] deals with the theory of absorption of low-frequency sound in the column itself and supports his conclusions by experiments. A loud-speaker "drives" a closed column 6·7 cm. in diameter, the amplitude and phase velocity in which are measured near the ends by manometric membranes. By varying the driving frequency, resonance curves for the column were obtained. The propagation of a rectangular wave-form of large amplitude (80,000 dyne/cm^2) was also examined. Even at these low frequencies carbon dioxide showed greater dissipation than air. On varying the mean pressure in the tube, the Helmholtz-Kirchhoff theory was verified, inasmuch as the damping was found to be inversely as the square root of the density. On stationary waves in tubes (Kundt's tube) we note a research by Robinson and Stephens[75]. By bubbling air through soap solution into a wide glass tube, a series of plane parallel soap films about 1 cm. apart was introduced. On exciting a loud speaker at one end at a suitable frequency, the films are set into membranous vibration, and further, if stationary waves are formed in the column, those at the places of maximum vibration are violently agitated and may be disrupted. Brandt and Freund[76] use a conventional Kundt's tube but provoke the dust figures with magneto-strictive sources of 6, 10 and 25 kc./sec. The usual patterns were produced, but of course on a very small scale, in liquids, tobacco smoke or other solid suspensions. In the latter case coagulation of the individual particles was brought about by the supersonic radiation. A comparison of the behaviour of hollow tubes filled with liquid and solid rods in respect of the propagation of supersonics has been made by Field[77], who draws attention to the importance of the shearing forces in the rod. An interesting application of the supersonic propagation in tubes has been made by Goldman[78] who constructed some model horns, 1–2 in. long and having various types of flare. He then let one end of a short steel rod, held in magneto-strictive oscillation, feed into their throats. In this ingenious way small-scale research was conducted into the directional characteristics of horns, using a microphone pick-up at various points round the source, in a small absorbent room. Polar curves, for a number of horns with or without baffles, at 25 kc./sec. were thus delineated.

A resistance thermometer for measuring the temperature in the interior of a sounding pipe is not unknown, but difficult to work, but for the first time a thermo-couple of sufficiently low thermal capacity has been constructed for this purpose. Johnson[79] used a cellulose acetate film spread over a mica frame 3 × 1 cm. having a central hole 5 mm. in diameter. This was "sputtered" from cathodes of antimony

and bismuth so that a covering of each metal was formed on either side, with a junction at the hole. The average thickness of the couple was $3{\cdot}5 \times 10^{-5}$ cm. but the mica frame holding the film was 0·1 mm. thick. Owing to the width of the frame it was not used in the middle of a pipe but at one end, a loud speaker driving the tube at the other. In calibration the couple was replaced by a condenser microphone. The thermocouple voltage is linear with sound pressure at constant frequency, but like all such instruments its sensitivity decreases with increase of frequency. Nevertheless the designer hopes to use it up to 300 kc./sec.

Marty[80] has been studying the variation of amplitude in the interior of various shapes of resonator by means of a Kundt's valve which allows only the compressions to pass to a manometer in which thereupon a static pressure will be built up, depending on the distance of the valve from a node. In place of the fly's wings used by a recent investigator, Marty prefers thin films of collodion for the valve.

In a short note Taber Jones[81] points out that a free reed may be maintained in oscillation over an orifice by a blast of air, as in the harmonium and mouth organ, in virtue of the inertia of the air. During vibration the flow of air is more rapid on the low pressure side of the tongue causing a greater reduction of pressure on that side and aiding the passage of the reed into its slot. When the air flow is cut off by the closing of the slot the reed moves out of the slot, but from the inertia of the air it results that the reed is not immediately driven back but continues in oscillation to and fro over the slot.

Redfield[82] raises some important points in connexion with the theory of wind instruments. The most telling of these is that the embouchure of, for example, the clarinet is not to be considered as a "stopped end" of the column of air in the instrument, for this column is effectively continued more or less into the player's mouth, larynx and chest. This explains the discrepancy between the theoretical speaking lengths—plus end-correction at the flare—and the actual frequency of such instruments. The mouthpiece end is in fact to be treated as approximating to an antinode, not as a node.

§ 4. ACOUSTIC IMPEDANCE AND PROPERTIES OF MATERIALS

There are several papers on filters to notice. Lindsay[83] together with Lewis and Albright[84] gives the theory of two new types of filter; one, consisting of a rod in torsional vibration loaded with equi-spaced discs, and the other, alternate strata of two different media. As we would expect, attenuation occurs in the latter filter when the width of either slice coincides with the wave-length of the oncoming vibration in this particular material due to resonant vibration, but frequencies are also rejected whenever the ratio of the specific impedances attains a certain value. The authors point out that although this type of filter may not be readily realized in practice, it may be regarded as the prototype of all reiterative conduits, since change of cross-section as well as change of density is equivalent to a change of specific impedance. This general formula for the attenuation band is

$$+1 \geqslant \cos 2k_1 l_1 . \cos 2k_2 l_2 - \tfrac{1}{2}\,(Z_2/Z_1 + Z_1/Z_2) \sin 2k_1 l_1 . \sin 2k_2 l_2 \geqslant -1.$$

Z, k and l being the respective impedances, wave-length constants and lengths of the alternate sections. This formula may quickly be adapted to practical filters by proper substitution for Z_1 and Z_2.

On the practical side Groszmann[85] has checked the dispersion and absorption theory of a low-pass filter in the form of an "infinite" glass flute, and high-pass filter in the form of a many-chimnied pipe, by a novel method, viz. by turning the filter into a Kundt's tube and comparing the wave-lengths of the dust figures for various frequencies with those in an unbroken tube of the same bore. So the modification of velocity in the filter is known. The absorption was measured in the conventional way with a microphone at the distant end of the filter.

G. D. West[86] has constructed a useful model for demonstrating the properties of filters. It consists of a series of equal masses m suspended on springs of restoring force k per unit displacement, distance a apart. The line consists of an elastic cord under tension F and threaded through holes in the masses. The input is furnished by an electric motor coupled to one end of the cord by a rocker arm which contributes a transverse simple harmonic motion to the system, which then acts as a band-pass filter whose limits are (*a*) the period of a single mass on its spring, and (*b*) the period which has a half wave-length equal to the distance between consecutive masses. The one limit corresponds then to resonance in the branch, the other to resonance in the line element of an acoustic (or electric) filter. Between these limits the length of the wave transmitted progressively changes, the relation between frequency and wave-length being

$$n=(2\pi)^{-1}\{(4F/ma)\sin^2\pi a/\lambda+k^2/m\}^{\frac{1}{2}},$$

which checks well with experiment. Considerable dispersion is shown by the filter, the wave velocity varying from 265 to 1080 cm./sec. within the range of the experiments.

There are two further pieces of apparatus for the measurement of impedance. Of Schuster's[87], we have few details at present beyond the fact that it is based on the Quincke interference tube. Sound from a telephone diaphragm is reflected by the two impedances to be compared at equal distances on either side of the telephone in a wide tube. Search tubes symmetrically placed with reference to the source lead off to a stethoscope arrangement. If the two impedances are equal no sound will be heard in the stethoscope but inequality of impedance will cause a sound which can be nullified by equalizing the impedances. At present the variable impedance consists of a stopped pipe of variable length. Measurements of impedance with this apparatus are promised later, though it is not clear how the value of the "standard" impedance is to be calculated.

Robinson's[88] apparatus is shown in figure 4. This apparatus is designed for the measurement of the "conductivities" of orifices. The source L is at one end of a tube, divided by the listening-tube T into two parts, l_1 to the left, l_2 to the right. The terminating impedance on the right (Z_2) is either the orifice C or the rigid stop. With the stop *in situ* so that $l_2=m\lambda/4$ where m is an odd integer, the impedance to the right is zero, and the sound in T a minimum. This sound may be further

reduced to zero by the adjustment of the impedance to the left, which is effected by altering the position of the piston in the side tube B. The unknown impedance C is then introduced in place of A and l_3 adjusted to restore silence. The impedance to the right is then equal to that of a stopped pipe of length l_3, viz. $(i\rho c/S)\tan kl_3$ and may be readily calculated. For the unknown impedance Robinson substituted circular orifices in metal plates, either forming constrictions in a stopped tube or debouching directly on the atmosphere. For small orifices (about 1 cm. in diameter) the classical formula

$$\text{conductivity} = \text{diameter}$$

was fulfilled, but conductivities as much as twice the diameter were observed when the diameter was increased to 4 cm.

The only other impedance paper is due to Littler[89] who has shown how to calculate the motional impedance of a telephone diaphragm by varying the pressure of the air load acting on it.

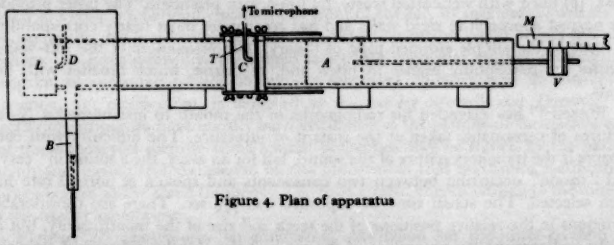


Figure 4. Plan of apparatus

In the domain of acoustic properties of materials we find measurements of the transmission of sound through plastic and amorphous substances by Gemant[90], and of materials common in India by Ghosh and Mohamed[91] while Chrisler[92] and Norris[93] are engaged in controversy as to how far the size of a specimen affects measurements of its reflection and transmission coefficients—an important consideration when small-scale methods are used. Spandock[94] has an interesting variant on the usual methods. Instead of allowing a continuous sound to fall on the specimen, he causes the source to emit a "short tone" of about $\frac{1}{100}$ sec. so that he may catch the whole of the reflected or transmitted train on the microphone after the train directly incident upon it has passed. Direct comparison of the amplitude of the two transmissions then gives the required coefficients.

§ 5. SUBJECTIVE ASPECTS OF SOUND

Dealing first with the phonetic side of speech, we remark that analysis of vowels and consonants continues to occupy a number of phoneticians and are pleased to notice a continued improvement, on the physical side, of the recording instruments used. Steinberg[95] prepares an oscillogram of a simple American phrase "Joe took father's show bench out", chosen apparently for the variety of sounds involved, and

analyses it from the aspect of continuity of wave-form throughout each vowel sound. He shows that the wave pattern in each vowel undergoes a slight but persistent change through some score of periods. For example in the sounds *ō*, *u* and *ū*, a resonance begins at about 1800 c./sec. and falls to 1200 c./sec. There may in addition be a change in the fundamental, i.e. the vocal-cord frequency, but this depends upon the training of the voice. From the rate of change of amplitude, damping coefficients of the order of 1000 decibels per second are suspected. This question of the damping of the vocal cavities, considered as soft-walled resonators, has also occupied Cotton[96]. The scheme involved obtaining the resonance curves for cylindrical vessels lined with various soft substances, and actuated by a constant-output telephone oscillator, adjusted to give the same ground level at all frequencies, as measured on a condenser microphone. The resonator was then coupled to the diaphragm and a new reading of the microphone taken. He obtains the resonance peaks of the resonator (14 cm. in length by 2 cm. in diameter), (*a*) lined with beef-steak, (*b*) lined with wet cotton waste, (*c*) lined with plasticine. The latter provides the nearest approach to rigid walls and has resonances most nearly corresponding to those of the simple stopped pipe of theory. The resonances of the soft-walled cavities are perceptibly higher in pitch and, of course, much broader with low crests.

Russell[97] has extended his radiographs of the mouth to instantaneous X-ray pictures of consonants taken at the instant of utterance. The difficulty with consonants is the transitory nature of the sound, but for an assay, the *z* sound in "easy" and "ozone" occurring between two consonants and spoken at normal rate has been selected. The actual time of exposure was $\frac{1}{1000}$ sec. There are considerable variations in the relative positions of the teeth and size of the mouth cavity, but in all the radiographs the separation of tongue and the alveolar ridge (just over the teeth of the upper jaw) is small, and suggests itself to Russell as the seat of the characteristic buzz of this consonant.

The other functional research on the voice is that of the writer[98], who has shown by laryngoscopic examination and the construction of hydrodynamic models that the false vocal cords, lying over the true ones, may under certain circumstances come into action to encourage the formation of an edge tone in the pulsating jet extruded from the true vocal cords.

Besides those of Steinberg there are oscillograms published by Obata and Tesima[99] (Chinese speech sounds), Thienhaus[100] (German vowels) and Curry[101] (Standard English and Northumbrian speech sounds).

Of papers which treat rather of the psychological aspects of voice production, we notice first some observations on the extent of the pitch and intensity range in ordinary speech and in declamation by Tiffin[102]. What he calls the "melody curves" of a radio saleswoman (a profession fortunately unknown in this country) and a Shakespearian actor have been delineated. He has endeavoured to correlate the range of syllabic intensities with the opinions of elocution experts listening to the recording, though up to the present without marked success. Evidently, as in judging other competitions, there is more in it than mere physical measurement!

Bartholemew[103] also endeavours to define in physical terms that elusive quality "good voice", this time in relation to singing. This quality seems to be characterized by (1) a formant of high frequency, in the neighbourhood of 3000 vibrations per second, having its origin in the larynx, (2) a formant of low frequency near 500 vibrations per second, (3) ease of production of intensity, and (4) a vibrato, not merely of pitch, but of intensity and timbre at the same time, having a period of about one-sixth of a second.

Finally Voelcker[104] has made a study of a type dear to psychologists, viz. the relative frequency of occurrence of speech sounds in formal American pronunciation, a study the importance of which the writer does not feel competent to assess.

On the subject of hearing, we find a number of papers dealing with non-linear distortion. While Braunmuhl[106] is concerned with the distortion of amplifying apparatus and its effect on the ear, von Bekesy[105], also Chapin and Firestone[108], are concerned with distortion in the ear itself. The former shows that overtones are added to a simple tone loudly sounded in front of the ear, by mechanical defects of the ear itself. He ascribes the seat of the distortion to the liquid which fills the cochlea. At great intensities this liquid is set into vorticular motion, causing the reception of incoming sounds to become selective. Another type of non-linear reception, i.e. rectification, is assigned to the ear by Stowell and Deming[107]. Rectifying properties have usually been attributed to the ossicles of the middle ear, but these observers seem to think similar properties may be shown by the basilar membrane, basing this conclusion on the observation that maximum rectification always occurs at the same frequency, i.e. 1000 c./sec., corresponding to the middle of the basilar membrane.

Fletcher[109] contributes an interesting discussion on the inter-relation of intensity, frequency and overtone-structure in respect to the sensations of loudness, pitch and timbre. He points out that none of these sensations is singular, that for instance the loudness scale of a pure tone is not the same as that of a complex note, and that a change of intensity with unchanged frequency may yet give the sensation of a change of pitch. He likens the mechanism of the ear to a set of tuning-fork resonators above 2000 c./sec., and a set of tuned strings below this frequency. At large displacements the restoring force in the former class is proportionally less, in the latter class it is proportionally greater. Thus the former decrease while the latter increase their resonant frequencies when unduly stimulated.

Vierling[113] treats of an analogous problem, that of the effect of intensity on the sensitivity of the ear to pitch, while Sell[110] applies similar considerations to the functioning of aids to the deaf. Churcher, King and Davies[111,112] explore by familiar methods the sensitivity of the ear to change of intensity, using a pure tone oscillator as source. As the sensitivity varies considerably with the original intensity level, it is concluded that the Weber-Fechner law does not apply. Geffcken[114] deduces what one might call the acoustic current in the ear from measurements of acoustic pressure at the outer ear (by thermophone) and of the acoustic impedance at the same place (for method cf. p. 90 of last *Report*). Sellerio[115] presents an interesting apparatus which might be called, after its optical analogue, a flicker

phonometer. The two sounds (presumably of equal pitch and timbre) to be compared fall severally on two microphones, coupled through transformers to a circuit containing a split resistance. The sliding contact which divides this resistance in two is connected to one side of a pair of headphones, the other side of which is connected alternately by a commutator to either end of the resistance. In general a "flicker" of intensity will be heard but this may be nullified by a suitable shift of the sliding contact. The ratio of intensities is then simply related to the ratio of the partial resistances. For objective measurements the headphones may be replaced by a vibration galvanometer.

A symposium held by the Bell Telephone Co.[116] discusses the problems set by the overland transmissions of music from one concert hall to another. The question is also discussed by Aigner and Strutt[117]. Although in theory an infinite number of transmission lines and units would be necessary to reproduce the music of an orchestra in true "auditory perspective", it generally suffices to have two microphones one each side of the stage at the sending end and two similarly placed loud-speakers in the reception hall. The faculty of binaural location does the rest. A concert of music from the Philadelphia Academy of Music was thus relayed recently to Washington, D.C.—a distance of 120 miles—for reproduction in auditory perspective, with gratifying results.

Although not directly concerned with hearing, we must notice here two papers of Goodfellow[118] and Gault[119] on the use of the sense of touch to replace that of hearing in the deaf. By experiments in which the subject touched a diaphragm vibrating at known frequency and amplitude, it was found that the minimum tactibility was some 40 to 100 decibels below minimum audibility at the same frequency. Few subjects however could detect vibrations above 4000 c./sec. From this to the ultimate aim of the investigators will be a long step, for they aim to train the deaf to interpret the sense of syllables transmitted to the diaphragm, by the feel of the timbre of its vibrations.

§ 6. TECHNOLOGY

As our allotted space is running out, we can only refer in detail in this section to those works of paramount novelty or importance. A number of investigators have been concerned with recording and intensity measuring apparatus. Thus Francke[120] has tested the thermophone and found that it does give a linear output as frequency is varied, in accordance with theory. Mallina[121] has devised a useful instrument for projecting a wave-form on a screen with sufficient intensity to be visible to an audience. A plan of the apparatus is shown in figure 5. The current from the microphone amplified and applied to an electromagnet near the mirror M causes the latter to oscillate about the axis AA. Light reflected from M falls on a rotating hexagonally shaped mirror N and is again reflected on to the screen SS (actually at a relatively greater distance than shown in the sketch). The arrangement causes a succession of pictures of the wave-form picked up by the microphone to pass across the screen, the mirrors being so disposed that one image enters the screen at one side as the preceding one leaves at the other, the time of traverse being approxi-

mately equal to that of the persistance of vision, viz. $\frac{1}{16}$ sec. Kluge[122] and Thienhaus[123] give some examples of timbre analysis in which the search-tone principle is employed. Suppose the complex note contains a component of frequency f_1; then, on blending with a pure sine wave in the analyser of frequency f_2, combinations of frequencies $f_1 \pm f_2$ are produced, whose amplitude is proportional to those of the reacting frequencies. Thus if f_2 has a fixed amplitude, that of f_1 can be measured by filtering out the difference tone and measuring its amplitude. In the present apparatus, the indicator tone is produced by a hexode oscillator.

Arkel and Urk[124] recommend the use of single crystals for gramophone needles as they find that the wear and tear on the disc is less than when the usual steel needle is used; further the scratch noise is less apparent.

Turning now to the subject of acoustics of buildings we have two theoretical papers to note. Faggiani[125] takes up the well-known formula of Sabine and re-

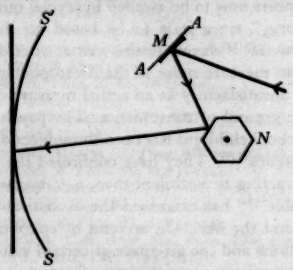


Figure 5

calculates it for the case of a room having a number of apertures such as we presume one might expect in a concert hall in a quiet neighbourhood on a hot day. The final form of the equation is not greatly affected. Brillouin[126], introducing a series of lectures on technical acoustics at the Collège de France, observed that there are a number of underlying assumptions in Sabine's work which have not yet received sufficient explanation to satisfy a physicist; among others, (*a*) that the time of reverberation must increase with the volume of the hall, (*b*) that this time should be reckoned from the cessation of the source, (*c*) that the rise and fall of intensity are exponential. He suggests that we should work in terms of a "purity coefficient" p defined as the ratio of the useful energy emitted in a syllable to the same energy plus the unwanted energy present at the same time, i.e. the residue of previous syllables. The latter he calculates in terms of the number of reflections per second, and consequently of the size of the hall, and finally obtains p in terms of the duration of the syllable and the average distances of the auditor and the walls respectively from the source. The increase of the optimum reverberation with size of hall then corresponds to the need for equal purity in all halls at points distant from the source. He then turns to the effect of reflectors in the hall and shows that such

surfaces are more efficacious for purity as they subtend a larger solid angle at the source, hence their current use behind a concert platform. He maintains, and it appears with reason, that the application of this new aspect in the theory will lead to a better understanding of practical problems in acoustics of buildings. On the experimental side we note a new departure due to Spandock[127], who studies such matters in models of auditoria about the size of large doll's houses, with enclosed source and pick-up. The use of such models raises of course questions of scale-effect which are not to be neglected. The present paper gives data on models of three typical shapes, the parallelopiped, the dome and the rotunda.*

On the topical subject of noise measurement and abatement we have a general account by Davis[131, 133] who *inter alia* stresses the need for a logarithmic unit in measurements of this type which shall be independent of a precise definition of a standard level (e.g. of minimum audibility), or of the validity of the Weber-Fechner law, which appears now to be assailed in several quarters. For this unit he proposes the name "brig" since it is to be based on the common logarithm. Churcher, King and Davies[132] revert to this matter after describing the methods which they have used to measure noise at the Metropolitan-Vickers works. They think the decibel scale unsatisfactory as an actual measure of sensation-level, even if it be retained on the ground of convenience. The psychological effect of noise on workers (Japanese school children) has been the subject of considerable research by Obata and his colleagues[135]. They have confirmed the generally held opinion that music is more distracting to a student than, e.g., typewriter noise. Finally, in this connexion, Constable[134] has examined the acoustical insulation afforded by double glass windows and the like. On account of coupling between the natural resonances of the partitions and the air-space at certain values of the separation of the panes, the actual transmission may at certain frequencies exceed that permitted by a single pane.

The inventors who would replace the organ by an all-electric musical instrument are still busy. Toulon[136] describes an interesting photoelectric organ of which the basis is a photoelectric projector similar to that used in cinemas, with stroboscopic discs replacing the moving film. By means of an ingenious system of slots and masks a variety of intensities and timbres may be simulated on any note. There is one disc to each octave, each disc having 13 rings of slots corresponding to the number of semitones in the diatonic octave. Chords and damped notes (pianoforte tone) may also be obtained. It has a thyatron (glow discharge) circuit, the grid-bias of which is altered by "stopping" a rheostat wire with a rubber finger, thus varying the frequency of oscillation.

Echo-sounding devices, too, continue to evolve. The latest use of this apparatus is in aircraft for enabling the pilot to determine his altitude and inclination to the ground. Delsasso[137] has tested such a contrivance on a small airship using the Doppler effect on the frequency of the emitted pulse for the soundings. Although

* Good general accounts of the present stage of research in this subject have been written by Knudsen[128] and P. Sabine[129] while Kaye[130] describes the new laboratory at the National Physical Laboratory.

the Doppler effect may fairly readily be used to measure the altitude, it appears that the inclination of a moving airship will not be easy to determine in practice, although covered by the theory.

REFERENCES

(1) Grossmann. *Phys. Z.* **35**, 83 (1934).
(2) Spakovski. *C.R. Acad. Sci. U.S.S.R.* **3**, 31 (1934).
(3) Richards and Reid. *J. chem. Phys.* **2**, 193, 206 (1934).
(4) Eucken and Becker. *Z. phys. Chem.* **27**, 219, 235 (1934).
(5) Kneser and Wallmann. *Naturwissenschaften*, **22**, 510 (1934).
(6) Curtis. *Phys. Rev.* **46**, 811 (1934).
(7) Rogers. *Phys. Rev.* **45**, 208 (1934).
(8) Rose. *J. chem. Phys.* **2**, 260 (1934).
(9) Richards. *J. chem. Phys.* **2**, 263 (1934).
(10) Huntington. *J. chem. Phys.* **2**, 441 (1934).
(11) Richardson. *Proc. roy. Soc.* **146**, 56 (1934).
(12) Sherratt and Griffiths. *Proc. roy. Soc.* **147**, 292 (1934).
(13) Keesom and van Lammeren. *Proc. Acad. Sci., Amst.* **37**, 614 (1934).
(14) Swanson. *J. chem. Phys.* **2**, 689 (1934).
(15) Hiedemann and Asbach. *Phys. Z.* **35**, 26 (1934).
(16) —— —— *Z. Phys.* **87**, 1442 (1934).
(17) Hiedemann, Asbach and Bachem. *Z. Phys.* **87**, 734 (1934).
(18) Bachem. *Z. Phys.* **87**, 738 (1934).
(19) Hiedemann, Bachem and Asbach. *Z. Phys.* **88**, 395 (1934).
(20) Hiedemann and Bachem. *Z. Phys.* **89**, 502 (1934).
(21) Hiedemann, Asbach and Hoersch. *Z. Phys.* **90**, 322 (1934).
(22) Hiedemann and Seifen. *Z. Phys.* **91**, 413 (1934).
(23) Hiedemann and Bachem. *Z. Phys.* **91**, 418 (1934).
(24) Debye, Sach and Coulon. *C.R. Acad. Sci.*, Paris, **198**, 922 (1934).
(25) Lucas. *C.R. Acad. Sci.*, Paris, **199**, 1107 (1934).
(26) Baumgardt. *C.R. Acad. Sci.*, Paris, **199**, 1383 (1934).
(27) Bär. *Helv. phys. Acta*, **6**, 570 (1934).
(28) Bär and Walti. *Helv. phys. Acta*, **7**, 658 (1934).
(29) Wyss. *Helv. phys. Acta*, **7**, 406 (1934).
(30) Sack. *Helv. phys. Acta*, **7**, 657 (1934).
(31) Claus. *Z. tech. Phys.* **15**, 74 (1934).
(32) Szalay. *Phys. Z.* **35**, 293 (1934).
(33) Liu and Wu. *J. Amer. chem. Soc.* **56**, 1005 (1934).
(34) Rocard. *Rev. Acoust.* **3**, 47, 192 (1934).
(35) King. *Canad. J. Res.* **11**, 135, 484 (1934).
(36) Skomtao and Su. *Nature*, Lond., **133**, 214 (1934).
(37) McFarlane. *Phil. Mag.* **18**, 24 (1934).
(38) Biot. *J. acoust. Soc. Amer.* **5**, 206 (1934).
(39) Hickman. *J. acoust. Soc. Amer.* **6**, 108 (1934).
(40) Meyer and Thienhaus. *Z. tech. Phys.* **15**, 630 (1934).
(41) Young. *J. acoust. Soc. Amer.* **6**, 112 (1934).
(42) King. *Proc. roy. Soc.* **147**, 212 (1934).
(43) Osterburg. *Rev. sci. Instrum.* **5**, 183 (1934).
(44) Bergmann. *Ann. Phys.*, Lpz., **21**, 553 (1934).
(45) Straubel. *Phys. Z.* **35**, 179 (1834); *Z. tech. Phys.* **15**, 607 (1934).
(46) Petrzilka and Zachoval. *Z. Phys.* **90**, 700 (1934).
(47) Schoeneck. *Z. Phys.* **92**, 390 (1934).
(48) Bechmann. *Z. Phys.* **91**, 670 (1934).
(49) Cremer. *S.B. preusz. Akad. Wiss.* **1**, 2 (1934).

(50) Schenk. *Z. Phys.* **88**, 626 (1934).
(51) Nemes. *Phil. Mag.* **18**, 303 (1934).
(52) Lark-Horovitz and Caldwell. *Nature*, Lond., **134**, 23 (1934).
(53) Hart, Fuller and Lusby. *J. acoust. Soc. Amer.* **6**, 80 (1934).
(54) Ghosh. *Phil. Mag.* **17**, 521 (1934).
(55) —— *Z. angew. Math. Mech.* **14**, 71 (1934).
(56) Davies and James. *Phil. Mag.* **18**, 1023, 1053 (1934).
(57) Dye and Essen. *Proc. roy. Soc.* **143**, 285 (1934).
(58) Araki. *Phil. Mag.* **18**, 441 (1934).
(59) Hacquebord and Huizing. *Beitr. Anat. etc., Ohr.*, **31**, 248 (1934).
(60) Brown. *J. Franklin Inst.* **218**, 41 (1934).
(61) Ruedy. *Canad. J. Res.* **10**, 134, 244 (1934).
(62) Raman. *Proc. Ind. Acad. Sci.* **1**, 179 (1934).
(63) Colwell and Bryant. *J. Franklin Inst.* **218**, 739 (1934).
(64) Obata and Tesima. *Proc. imp. Acad. Japan*, **10**, 211 (1934).
(65) Pomp and Zapp. *Engineer*, **157**, 120 (1934).
(66) Curtiss and Giannini. *J. acoust. Soc. Amer.* **5**, 213 (1934).
(67) Waller. *Proc. phys. Soc.* **46**, 116, 124 (1934).
(68) Carrière. *Rev. Acoust.* **3**, 221 (1934).
(69) Zickendraht. *Helv. phys. Acta*, **7**, 773 (1934).
(70) Harding. *Helv. phys. Acta*, **7**, 655, 834 (1934).
(71) Meier. *Phys. Z.* **35**, 524 (1934).
(72) Hopfield. *Nature*, Lond., **134**, 737 (1934).
(73) Press. *Phys. Z. Sowjet.* **5**, 616 (1934).
(74) Lehmann. *Ann. Phys.*, Lpz., **21**, 533 (1934).
(75) Robinson and Stephens. *Phil. Mag.* **17**, 27 (1934).
(76) Brandt and Freund. *Z. Phys.* **92**, 385 (1934).
(77) Field. *Canad. J. Res.* **11**, 254 (1934).
(78) Goldman. *J. acoust. Soc. Amer.* **5**, 181 (1934).
(79) Johnson. *Phys. Rev.* **45**, 641 (1934).
(80) Marty. *Ann. Phys.*, Paris, **1**, 622 (1934).
(81) Jones. *Rev. sci. Instrum.* **5**, 192 (1934).
(82) Redfield. *J. acoust. Soc. Amer.* **6**, 34 (1934).
(83) Lindsay. *J. acoust. Soc. Amer.* **5**, 196 (1934).
(84) Lindsay, Lewis and Albright. *J. acoust. Soc. Amer.* **5**, 212 (1934).
(85) Groszmann. *Ann. Phys.*, Lpz., **21**, 433 (1934).
(86) West. *Proc. phys. Soc.* **46**, 186 (1934).
(87) Schuster. *Phys. Z.* **35**, 408 (1934).
(88) Robinson. *Proc. phys. Soc.* **46**, 772 (1934).
(89) Littler. *J. acoust. Soc. Amer.* **5**, 235 (1934).
(90) Gemant. *Phys. Z.* **35**, 167; *Z. Phys.* **87**, 700 (1934).
(91) Ghosh and Mohamed. *Indian J. Phys.* **9**, 167 (1934).
(92) Chrisler. *J. acoust. Soc. Amer.* **6**, 115 (1934); *Bur. Stand. J. Res.*, Wash., **13**, 169 (1934).
(93) Norris. *J. acoust. Soc. Amer.* **6**, 43 (1934).
(94) Spandock. *Ann. Phys.*, Lpz., **20**, 328 (1934).
(95) Steinberg. *J. acoust. Soc. Amer.* **6**, 16 (1934).
(96) Cotton. *J. acoust. Soc. Amer.* **5**, 208 (1934).
(97) Russell. *J. acoust. Soc. Amer.* **5**, 247 (1934).
(98) Richardson. *Proc. Univ. Durham phil. Soc.* **9**, 146 (1934).
(99) Obata and Tesima. *Proc. imp. Acad. Japan*, **10**, 322 (1934).
(100) Thienhaus. *Z. tech. Phys.* **15**, 637 (1934).
(101) Curry. *Proc. Univ. Durham phil. Soc.* **9**, 153 (1934); *J. sci. Instrum.* **11**, 162 (1934).
(102) Tiffin. *J. acoust. Soc. Amer.* **5**, 225 (1934).
(103) Bartholemew. *J. acoust. Soc. Amer.* **6**, 25 (1934).
(104) Voelcker. *J. acoust. Soc. Amer.* **5**, 242 (1934).
(105) v. Bekesy. *Ann. Phys.*, Lpz., **20**, 809 (1934).

(106) Braunmuhl. *Z. tech. Phys.* **15**, 607 (1934).
(107) Stowell and Deming. *J. acoust. Soc. Amer.* **6**, 70 (1934).
(108) Chapin and Firestone. *J. acoust. Soc. Amer.* **5**, 173 (1934).
(109) Fletcher. *J. acoust. Soc. Amer.* **6**, 59 (1934).
(110) Sell. *Z. tech. Phys.* **15**, 30 (1934).
(111) Churcher, King and Davies. *Phil. Mag.* **18**, 927 (1934).
(112) Davies. *Phil. Mag.* **18**, 940 (1934).
(113) Vierling. *Z. tech. Phys.* **15**, 641 (1934).
(114) Geffcken. *Ann. Phys.*, Lpz., **19**, 829 (1934).
(115) Sellerio. *Nuovo Cim.* **11**, 587 (1934).
(116) Fletcher and others. *Bell. Syst. tech. J.* **13**, 239 (1934).
(117) Aigner and Strutt. *Z. tech. Phys.* **15**, 355 (1934).
(118) Goodfellow. *J. acoust. Soc. Amer.* **6**, 45 (1934).
(119) Gault. *J. acoust. Soc. Amer.* **5**, 252 (1934).
(120) Francke. *Ann. Phys.*, Lpz., **20**, 780 (1934).
(121) Mallina. *J. acoust. Soc. Amer.* **6**, 37 (1934).
(122) Kluge. *Z. tech. Phys.* **15**, 223 (1934).
(123) Thienhaus. *Z. tech. Phys.* **15**, 637 (1934).
(124) Arkel and Urk. *Physica*, 's Grav., **1**, 425 (1934).
(125) Faggiani. *RC. Accad. Lincei*, **19**, 887 (1934).
(126) Brillouin. *Rev. Acoust.* **3**, 166 (1934).
(127) Spandock. *Ann. Phys.*, Lpz., **20**, 345 (1934).
(128) Knudsen. *Rev. mod. Phys.* **6**, 1 (1934).
(129) Sabine. *J. Franklin Inst.* **217**, 443 (1934).
(130) Kaye. *Nature*, Lond., **134**, 202 (1934).
(131) Davis. *Engineering*, **138**, 663 (1934).
(132) Churcher, King and Davies. *J. Instn elect. Engrs*, **75**, 401 (1934).
(133) Davis. *Proc. phys. Soc.* **46**, 631 (1934).
(134) Constable. *Phil. Mag.* **18**, 321 (1934).
(135) Obata and others. *J. acoust. Soc. Amer.* **5**, 255 (1934).
(136) Toulon. *Rev. Acoust.* **3**, 83 (1934).
(137) Delsasso. *J. acoust. Soc. Amer.* **6**, 1 (1934).

HEAT

By J. H. AWBERY, B.A., B.Sc., F.Inst.P.,
National Physical Laboratory

§ 1. TEMPERATURES (ATTAINMENT, MAINTENANCE AND MEASUREMENT)

In the report on heat published last year, it was recorded that the demagnetization technique had enabled a temperature of 0·085° K. to be reached. This achievement was almost immediately surpassed, the lowest figure recorded during 1934 showing that five-sixths of the remaining distance to the absolute zero had been covered. It is also to be remarked that despite the difficulty referred to in that report, respecting the transfer of heat from the demagnetized body to any other, it has been possible to discover several new supraconductors in the region now made accessible[1]. For experiments at these temperatures, the demagnetized body and the substance to be investigated are previously made into a pellet, so that satisfactory cooling is achieved.

A book[2] on low temperature physics has recently been published and is very useful, although unfortunately it appeared too early to give the latest results obtained by the demagnetization technique.

From the point of view of the establishment of a uniform scale for low temperatures, the data obtained in different laboratories for the boiling-point of oxygen are of interest. This point is the lowest at present included in the International Temperature Scale and is therefore particularly important on this account. After correction of the older values by the use of a more accurate figure for the expansion coefficient of the glass employed, the three determinations from Leiden are:

1908	...	−182·97° C.
1918	...	−182·954° C.
1930	...	−182·977° C.

The authors of the latest of these determinations, Keesom, Miss van der Horst and Miss Jansen[3], adopt a weighted mean of −182·97° C.

The same quantity has twice been measured at the Reichsanstalt, the 1914 value being −182·97, and the 1924 one −183·00° C. There is thus a discrepancy of 0·03° C. between the results arrived at by the two laboratories, an amount which is greater than the combined errors as estimated by the investigators.

At a much lower temperature, the data for the boiling-point of hydrogen are of interest[4, 5]. The question is complicated by the fact that this gas can exist in two forms, ortho- and para-hydrogen. Normal hydrogen contains three volumes of the former to one of the latter, though at low temperature the proportions can

be altered artificially by absorption. The results (for normal hydrogen) are shown below:

Observers	Laboratory	Date	Method	Result	Result corrected by Keesom (1932)
K. Onnes and Keesom	Leiden	—	Pt thermometer	−252·76	−252·754
C. Cath	Leiden	—	He thermometer	−252·70	−252·729
Palacios and K. Onnes	Leiden	—	He thermometer	−252·74	−252·769
Keesom, A. Bijl and H. v. d. Horst	Leiden	1932	He thermometer	−252·754	−252·754
Henning	Berlin	1921	Pt thermometer	−252·80	—
Henning and Heuse	Berlin	1924	He thermometer	−252·78	—
Heuse and Otto	Berlin	1931	He thermometer	−252·780	—

At still lower temperatures, where the vapour pressure of helium is the thermometric standard, the Leiden Laboratory has used three scales at different dates, and a useful comparison of these is given by Keesom[6].

Perhaps the most fundamental point of all, the ice-point, also continues to receive attention; when high precision is desired, the purity of the ice used is of importance, and even the question whether it contains air must be considered. Again, since the melting-point of ice varies with pressure, it has at times been suggested that the triple point (where the applied pressure is uniquely fixed, being the pressure of vapour in equilibrium with ice and water simultaneously) should be used. During the year under review, W. P. White[7] has shown that it is in practice less reproducible than the ice-point under atmospheric pressure, which can be reproduced to within 0·0001° C. It is to be noticed that the two questions are related, since when the pressure is lowered in order to reach the triple point, any dissolved air is unavoidably released.

In White's view, the best method of obtaining a steady temperature of accurately zero, is to use an ice jacket around the ice and water-bath proper, and to wash the ice itself as often as may be necessary to obtain the desired reproducibility. The question has also been examined by Thomas[8], who reached very similar conclusions. He points out that the ice must be pure, and that impurities may be gaseous as well as solid. He warns workers against having open flames or other sources of contaminating gases in the same room as an ice-point apparatus.

As to the position of this point on the thermodynamic scale of temperature, the most recent result seems to be that of the Leiden workers. Keesom and van Santen measured corresponding values of pressure and specific volume for helium at temperatures of 0°, 20° and 100° C. and found that pv at each temperature satisfies an equation of the form $pv = A(1 + B/v)$, where A and B vary with temperature. From the point of view of thermometry, the important quantity is $B_{100} - B_0$, for which they obtain $-0{\cdot}020_1 \times 10^{-3}$, previous observers having found values from 0·016 to 0·020$_8$. When their value is used with the Leiden determination of the pressure coefficient of the helium thermometer, they find the ice-point to lie at 273·14° above absolute zero.

Since the date of this work, Keesom, Miss van der Horst and Taconis[9] have redetermined the fundamental pressure coefficient of helium. Their new value does not affect the calculated figure for the ice-point. In fact, Heuse and Otto[10] (1929) in Germany, Jacobus[11] (1930) in America, and the above authors[9] (1934) in Holland, have given the values 273·16, 273·16 and 273·14° K., so that the true value can hardly differ from 273·15 by more than a hundredth of a degree.

The other fundamental point (the steam-point) is subject to much greater fluctuations with pressure, and a new determination by Moser[12] gave the result that the boiling-point of water on the International scale, at a pressure p mm. of mercury (between 680 and 780) is given by the formula $100+0{\cdot}03867\,(p-760)-0{\cdot}000022\,(p-760)^2$. The comparison between this formula and earlier results is shown in the table below.

p (mm.)	Boiling-point (° C.)		
	Chappuis (recalculated by Volet, 1929)	Zmaczynski, 1930	Moser, 1932
680	96·904	96·909	96·910
700	97·706	97·709	97·709
720	98·488	98·488	98·490
740	99·252	99·252	99·254
760	100·000	100·000	100·000
780	100·732	100·731	100·729

An insidious error arises when mercury or other thermometers are used for measuring the temperature of a gas stream, and is due to the frictional heating of the thermometer itself by the gas. Brun[13] has examined experimentally what is the temperature rise when a cylindrical object moves transversely through air or carbon dioxide with velocity V cm./sec., and finds that his results are represented by the formula $\theta=\alpha V^2$, where θ is the temperature rise in ° C. For air, $\alpha=4{\cdot}8\times10^{-8}$ whilst for carbon dioxide it is $5{\cdot}9\times10^{-8}$. He has also examined the variations in temperature around the circumference of the cylinder[14], though these figures are of little interest, since they must depend so greatly on the thermal conductivity of the material from which it is constructed.

For use at high temperatures, Zetzmann[15] has urged the advantages of pyrometers incorporating photoelectric cells.

A very thorough investigation was carried out by Haase[16] before designing his colour pyrometer. The principle of this instrument is the matching, not of intensities as in the usual optical pyrometer, but of actual colour. The light from the radiating body is split into beams, which are passed through different coloured filters, and then mixed in controlled proportions until they match a standard colour. To see how this gives a measure of the temperature, imagine that a match has been obtained, and that the temperature is then raised. Not only will the intensities of the beams coming through both filters be increased, but that of the short-wave colour will be increased relatively to the other; thus the mixture will be bluer than

before, and a match can be restored by weakening the corresponding beam. The degree of weakening is clearly a measure of the change in temperature.

The method has the advantage that for grey bodies (those for which the emissivity relative to a black body is independent of wave-length), the colour temperature is equal to the black-body temperature, and such important materials as iron, nickel, aluminium, tungsten and molybdenum are practically grey.

In the measurement of flame temperatures, the spectral line reversal method has been subjected to criticism by David[17] on several occasions. He claims that it shows higher temperatures than the calculated "theoretical" temperatures, which must themselves be upper limits; if this is so, it is conclusive proof that the spectral line method does not give the temperature corresponding to the mean kinetic energy of the gas molecules, and hence that this energy and the excitation (electronic) energy of the sodium atom do not reach equilibrium under the circumstances of his experiments. Clearly this is a promising field for further research, both on the experimental and theoretical sides. It may be that the explanation lies merely in the fact that the spectral line method gives the maximum temperature, which may without inconsistency be higher than the average temperature over the cross-section, as calculated from theory. Indeed, a paper by Ellis and Morgan[18] on the temperature gradients in flames, shows that differences of the order of 200° C. can occur in explosive flames, and this would possibly be sufficient to explain the discrepancies noted by David. The fact that the spectral line reversal method gives the correct value for a mass of gas heated in a furnace, shows that whatever may be the case in a flame, equilibrium is ultimately reached between the two forms of energy when sufficient time is given. On the other hand, if lack of equilibrium is the cause, it is difficult to see why Griffiths and the writer[19] should have found agreement between the spectral line method and that depending on the heat loss from a wire carrying current, when the conditions were those of an ordinary flame.

Turning to measurements of particular temperatures, there is little to record during the period under review, save the determinations of a number of boiling- and melting-points which can usefully serve as standardizing points. The close agreement between the determinations of the freezing-point of platinum at the National Physical Laboratory and the German Reichsanstalt, which was mentioned last year, would appear not to have been accidental: the American National Bureau of Standards has now published a value which lies within $\frac{1}{4}$° C. of both the others. It is also of interest to notice that a determination of the melting-point of rhodium (1966° C.) has been carried out in America by direct comparison[20] of the intensity of the light from gold and from rhodium, each at its melting-point. The value obtained is believed to be correct within 3° C.

For the production of very high temperatures, practically the only resistance elements sufficiently infusible are tungsten and carbon. A particularly well-designed furnace using the former material was constructed some time ago at the German Reichsanstalt. A number of authors have described carbon resistance furnaces, one of the most recent and most interesting being that of R. D. Evans[21]. This author uses a vertical graphite tube, split down a diameter along all but the last inch of

its length. The current passes down one half, then across the uncut piece, and returns along the other half, the two being separated at the top by an insulating distance piece. One advantage of this construction is that the tube is not constrained axially, and is thus less liable to crack when it expands. Another is that only one end needs to be water-cooled, and consequently there is a greater length of uniform temperature. With the insulation he has used, Evans finds that the furnace will reach 2000° C. with a power input of 7 kw.

The need to maintain a constant temperature is one which is encountered at some time or another in most branches of physics, and the authors who have devised and described thermostats are legion. In many cases the relay as well as the sensitive mechanism is the subject of special design.

When the heating medium is gas, the favourite method is to use the expansion of some liquid such as toluol to regulate the supply of gas, generally with mercury to transmit the alterations in level to the operative mechanism. Another method is to allow the curvature of a bimetallic strip when heated, to perform the same function.

When the heating is electrical, a thermocouple is frequently used as the sensitive element, its increase of e.m.f. when the temperature rises being used to operate a relay which switches in resistance and so decreases the heating current. In the same way, a resistance thermometer may be used; it has in fact been proposed for high temperature control, that the windings of the furnace should be used for this purpose. The method then has the advantage that fluctuations in temperature begin to be corrected at the very first instant that they are perceptible, and certainly before they become observable inside the furnace.

Outfits for electrical thermostatic control have been described by Lange and Voos[22], who use a photoelectric cell and thyratron to actuate the relay, the cell being brought into action by the light reflected from the mirror of a galvanometer connected to a thermocouple in the furnace. They claim that a furnace at 1000° C. can be maintained at a temperature which varies appreciably less than 0·2° C. over long periods of time.

For a totally different range, viz. that between −70° C. and 0° C., Cameron[23] has described a thermostat which keeps constancy within 0·1° C. for hours. The regulator is of the usual toluene and mercury type, the expansion of the toluene when the temperature rises causing the mercury to complete a circuit through platinum wires sealed through a glass tube. When the circuit is closed, a valve admits compressed air into a reservoir of acetone cooled by solid carbon dioxide, and thus forces this cold liquid into the bath under control. When the contact is open, liquid can return to the reservoir by a siphon action. Stirring is doubtless an extremely important matter with such a method as this, since the newly admitted liquid must be well mixed with that already in the bath, if uniformity of temperature is to be obtained. Over the same temperature range, the cryostat described by Schattenstein[24] gives even closer control. In this type, a copper rod with one end in a reservoir of solid CO_2 projects into the bath to be controlled, thus tending to cool it continuously. The rate of extraction of heat can be varied by adjusting the

length of, and therefore the temperature gradient in, the rod. The rate of cooling is arranged to be slightly greater than the natural rate of leakage of heat into the bath, so that the temperature tends to fall, and this tendency is overcome by electrical heating controlled in the usual way by means of a toluene regulator and relay.

§ 2. CALORIMETRY

The need of the biologist to know the energy changes in vital processes has been the cause of a number of papers on microcalorimetry, i.e. the measurement of very small quantities of heat, or of very slow rates of heat evolution.

Schukarew, Krivobabko and Schukarewa[28] have used a change-of-state calorimeter for this purpose. As in the Bunsen ice calorimeter, they utilize the volume change on melting to give indications of the quantity melted, but in order to work at normal temperatures they replace the ice by diphenylmethane. The choice of this substance appears to have been the result of a good deal of exploratory work.

Based on quite a different principle is the apparatus of Swietoslawski[29] and the more recent variation of it designed by Sanielevici[30]. The calorimeter is a metal block of very small dimensions, and is surrounded by a jacket maintained always at the same temperature as the calorimeter, so that there is no heat loss from the latter however long the experiment continues. The second mentioned author has used the apparatus for measuring the heat evolution of radioactive salts.

The measurement of the heat emitted on the absorption of the γ-rays from radium is akin to those just dealt with. Zlotowski[31] has measured the temperature rise in lead surrounding the radium source. He used a spherical vessel, and could vary the thickness by adding successive shells of lead until a constant condition was reached. His preliminary result is $8{\cdot}9 \pm 0{\cdot}3$ cal. per hour per gram of radium.

For less specialized purposes, Tian[32] has devised a most ingenious micro-calorimeter, in which the temperature is maintained absolutely constant, the heat emitted in the reaction concerned being compensated by the Peltier effect between two dissimilar metals. In such a case, the heat capacity of the calorimeter does not enter into the problem, and need not be known. Moreover, the problem of jacketing becomes simply that of maintaining a constant temperature in the surrounding vessel.

In the view of this author, it is so important to avoid heat production due to the stirrer, that he dispenses with it, making the apparatus so small that uniformity of temperature is attained without any stirring. Again, since the Peltier effect is only sensible with currents so small that the Joule effect is negligible, the calorimeter is only adapted for dealing with extremely small evolutions of heat. Thus, the production of 0·001 gm./cal. in 7 c.c. of water can be measured within a few per cent with the apparatus designed by Tian; in this case, instead of compensating the effect, the out-of-balance e.m.f. is observed directly. Examples of determinations carried out with the apparatus include the determination of the heat given out by flies, the heat of dilution of potassium nitrate (using only 0·8 c.c. of added water), and the heat of absorption of hydrogen on copper.

Another calorimeter in which the heat is measured by the temperature rise in a block of material, is that due to Whipp[33]. Contrary to the tendency exhibited by most designers, he uses only five thermocouples to measure the temperature change, but obtains high sensitivity (3×10^{-6}° C. per mm.) by keeping the resistance low. With a water equivalent of 1 gm. he thus secures a thermal sensitivity of about 3×10^{-6} cal. per mm. deflection. The effect of random errors is largely reduced by the use of a second, similar calorimeter, which acts as a dummy, and is placed symmetrically to the real calorimeter, within the same cylindrical enclosure.

The same author, working with Roberts, also carried out an interesting measurement of the heat of absorption of hydrogen on tungsten[34], the temperature rise of the latter being deduced from its resistance, when hydrogen was admitted to the previously evacuated apparatus.

Since water is still the commonest calorimetric medium, the question of the variation of its specific heat with temperature is an important one; in the absence of knowledge of this variation, results obtained at a mean temperature other than 15·5° C. could not be expressed in terms of the standard calorie. A recent investigation by Jessel[35] has led him to claim that the curve of specific heat against temperature for air-free water differs greatly from that for air-saturated water.

An interesting piece of calorimetric work is that of Kraus and Ridderhof[36], who have measured the heats of solution of various substances in liquid ammonia. As a measure of the heat evolved, they observe the quantity of liquid ammonia contained in an outer bath which is evaporated by the reaction concerned. The latent heat of evaporation of ammonia being one of the most accurately known for a very wide range of pressures and temperatures, they consequently obtain results of considerable precision, which would be difficult to attain in any other way.

Stephens[37] has used a cooling method in which the main source of heat loss is not the radiation and convection from the walls, but takes place through a piece of poorly conducting material introduced for this purpose. The instrument must be calibrated with liquids of known specific heat, but is then probably very convenient for many purposes.

Among more orthodox calorimeters, may be mentioned that of Askew and his collaborators[38], which uses water as calorimetric medium.

Remembering that the specific heat at constant volume is the temperature-derivative of the internal energy of a substance, it is obvious that it must provide a sensitive indication of any changes which affect the internal energy. Thus, in the demagnetization method of obtaining low temperatures, the rate of cooling of a pellet consisting of the demagnetized salt together with some other substance immediately gives notice of the occurrence of any change of this type. Again, the abnormal dielectric constant—20,000 c.g.s. units—of Rochelle salt, led Rusterholz to examine its specific heat; he finds that above 23° C. it rises some 5 per cent, falling to the normal value again at about 26° C. It will be remembered that Simon discovered the transformation which occurs in ammonium chloride at about −30° C. in a precisely similar manner.

The determination of heats of combustion of gases being a process of frequent

industrial occurrence, a special technique has been developed to deal with it. It is customary, for example, to determine the water equivalent of the calorimeter by combustions of standard substances, such as benzoic acid, naphthalene and succinic acid. Consequently, absolute measurements of their heats of combustion have a special significance.

Keffler[39] has determined the heat of combustion of succinic acid, which he finds to be sufficiently easily ignitable, by means of an adiabatic calorimeter. His final series of five experiments shows no variations greater than one part in 3000, the mean value at 20° C. being 3025·3 cal. per gm. This may be compared with the older determination of Verkade and Coops, who found 3026·9 cal. per gm.

Similarly, Jessup and Green[40] have studied benzoic acid anew. Their calorimeter was electrically standardized, and they express their result therefore in electrical energy units.

The mean results of a very large number of experiments on two samples were 26,436 and 26,434 international joules per gm. (weighed in air) or 26,414 weighed *in vacuo*. They correct the values found by previous observers to a standard temperature and set of energy units, as well as to a standard initial state, thus finding the variation in six determinations from 1909 to 1928 to be only from 26,415 to 26,437.

An unusual problem in calorimetry is the determination of the quantity of energy rendered latent when metals are hammered, rolled, drawn or otherwise cold-worked. Taylor and Quinney[41] have measured this by direct thermal methods (*a*) measuring the temperature rise of the specimen of known heat capacity, and (*b*) dropping it into a calorimeter so as to measure the heat directly. The two methods agreed very closely. They find that the work necessary to saturate copper with latent energy is about 14 cal. per gm., of which a great part can be recovered.

France[42] has adopted the method of measuring the heats of solution of the unstrained and of the cold-worked metal in a suitable solvent. The latent energy is then revealed as an increase in the heat of solution. By this method he found that drawn copper wire after a reduction of area of 95 per cent had absorbed less than 0·5 cal. per gm., whilst for cold-rolled iron he found rather larger figures—up to 7·3 cal. per gm. for 94 per cent reduction in area.

The inter-relations between the different branches of physics are well illustrated by considering with these the work of Wood[43], who from measurements of the broadening of the X-ray lines in cold worked metals, has deduced the alteration in lattice spacing and hence the change in potential energy. By this radically different method, he finds the latent energy in copper twisted to maximum lattice distortion to be 1·7 cal. per gm.

§ 3. CONDUCTION

The thermal conductivity of a gas has always been a difficult quantity to measure, and it is not surprising that a number of workers have devoted their energies to overcoming the difficulties. To take the case of the commonest gas, air, we find that at the date of preparation of International Critical Tables (about 1929)

the best value was believed to be $5{\cdot}33 \times 10^{-5}$, whereas now we believe (on much stronger evidence) that it is close to $5{\cdot}77 \times 10^{-5}$.

The history of the approach to truth in this matter is an interesting chapter. At the time when the Tables were prepared, it seemed that the wire method, in which the radial flow of heat is used, gave higher values than the parallel plate method, as seen in the table below:

Date	Observer	$10^5 \times$ Conductivity of air (0° C.)	
		Wire method	Plate method
1927	Weber	5·74	—
1909	Todd	—	5·71†
1926	Gregory and Archer	5·83*	—
1926	Schneider	5·88	—
1919	Hercus and Laby	—	5·40
	Means	5·82	< 5·4

* Raised to 5·85 in 1933. † At 55° C.

In principle, the plate method should give the more accurate result, since in it the temperature discontinuity at the air-solid interface is less important, and it is probably easier to eliminate convection in this form of apparatus.

Quite recently, Kannuluik and Martin[45] working in Laby's laboratory have given the results of determinations by a hot-wire method, in which both the above difficulties have been very specially attended to. They prefer to use relatively short and stout wires, which have many advantages over the long thin ones usually chosen in earlier work. On the other hand it is impossible to have any appreciable length in which the temperature is really uniform, so that a much more accurate theory of the temperature distribution in the wire and in the gas must be employed. The result which they reach is $5{\cdot}76 \times 10^{-5}$ c.g.s. unit.

Dickins[46] has also made use of the hot-wire method, employing it in such a form that the accommodation coefficient for the energy transfer when gas molecules strike the solid surface can be determined at the same time as the thermal conductivity. The experimental work appears to be of very high accuracy, and the result for air at 0° C. is $5{\cdot}84 \times 10^{-5}$ c.g.s. unit.

It may also be of interest to quote the values found for the accommodation coefficient in this work, in comparison with those found twenty years ago by Knudsen, from experiments in rarefied gases.

Accommodation coefficients between platinum and various gases

	Knudsen	Dickins
Helium	0·34	0·50
Argon	0·86	0·89
Hydrogen	0·36 (0·28)	0·36
Oxygen	0·84	0·90
Nitrogen	0·87	0·89
Nitrous oxide	0·73	0·89
Carbon dioxide	0·86	0·91

Again, it is to be noted that Gregory and Archer have rediscussed[47] their earlier results, and now assign the value 5·85, whilst Milverton[48] has obtained the value 5·81, also by the wire method. Both these values are corrected for the effect of accommodation. The recent values by this method are thus

Date	Observer	$10^5 \times$ conductivity of air
1934	Kannuluik and Martin	5·76
1934	Dickins	5·84
1934	Milverton	5·81
1933	Gregory and Archer	5·85
	Mean	5·82

In the same period, Hercus and Sutherland have carried out an extremely careful redetermination[49] by the parallel plate method. They consider that convection was not present, whilst radiation was eliminated by working at two different distances. The temperature differences varied from 10 to 16° C., and the plates had an area of 332 sq. cm. The result reduced to 0° C. was 5·72, so that although the plate method still gives lower results than the hot-wire method, the difference has been markedly reduced, and it is very improbable that the value can differ from $5{\cdot}77 \times 10^{-5}$ by more than 1 per cent.

Turning now to a totally different branch of the subject, we find that the thermal conductivity of metals has received a good deal of attention. R. W. Powell[50] has published the results of an absolute determination up to 800° C. for iron of a high degree of purity. In his method, the heat flows longitudinally along a bar of the material under test, which is jacketed by a tube in which a temperature gradient similar to that in the bar itself is maintained. Thus to a first approximation there is no edge loss. He measures the heat-flow producing the observed temperature gradient by electrical measurements of the energy dissipated in the heater, and avoids an obvious source of doubt by using two similar rods with the heater between them.

A somewhat analogous method has been used by Van Dusen and Shelton[51]. They also employ rods of metal, with a guard tube which eliminates lateral heat loss, but they measure the heat flow by observing the temperature gradient in a comparison bar of metal placed end to end with the test specimen. As primary standard, they adopt lead, and assume that its thermal conductivity at 0° C. is known. Its variation with temperature can be determined in the apparatus itself. As a more convenient secondary standard, they use nickel, which was first compared directly with lead. The consistency of the results was checked by interchanging the lead and nickel, and also by determining directly the temperature coefficient of the conductivity when both bars were of nickel.

This apparatus was used by Shelton to measure the conductivities of iron and a number of its alloys. It is interesting to compare his results for the purer irons with those obtained by Powell and others.

Thermal conductivity of pure iron

Purity (%)	Observer	Temperatures (° C.)				
		100	200	300	400	500
100·00	Powell	0·176	—	—	—	—
100·00	Masumoto	0·168	—	—	—	—
99·92	Powell	0·163	0·147	0·132	0·116	0·104
99·90	Shelton	0·159	0·145	0·131	0·118	0·104
99·49	,,	0·141	0·130	0·119	0·108	0·097
98·71	,,	0·110	0·104	0·099	0·093	0·088

The question of the variation of the Lorenz function $\kappa/\lambda T$, where κ and λ are the thermal and electrical conductivities, is still rousing considerable interest. The Sommerfeld theory of electrons in metals, like the earlier theory of Drude, predicts that this function should be constant, but actually it is found to vary very considerably with temperature. Grüneisen and Reddeman[52] have examined the matter for alloys of several metals. They deal not with conductivities but with resistivities, so that if ρ is the electrical and H the thermal resistance, the theoretical relation would be $\rho/HT = \text{constant}$, and a graph of ρ/T against H would give a straight line passing through the origin. Actually their results in general give a straight line intercepting the axis $\rho/T = 0$ at a finite value of H, when the results for a given series of alloys at a fixed temperature are plotted. As an example, they studied Ag-Au alloys varying from 0 to 75 per cent Au, and including three samples of pure silver, two of which were single crystals and one polycrystalline, and found a linear relation for the whole range at $-190°$ C., though not at $-251°$ C. If alloys of one main constituent with *different* metals are considered, then it is found that for a certain range of concentrations, the points all lie on the same straight line, i.e. its slope and intercept are characteristic of the main metal.

The results are interpreted as showing the effect of lattice conduction superposed on the electronic conduction. The authors mentioned deduce that the linear relation holds up to a definite value of the lattice distortion, the limiting distortion decreasing with decreasing temperature. Outside the range of validity of the linear relationship, it is found that atoms which are equivalent as regards their effect on electrical conductivity, are not so for thermal conductivity. Consequently the Norbury rule, according to which the effect of a substituted atom increases with its horizontal separation from the main metal in the periodic table, cannot be extended to rich alloys at low temperatures. The rule is, however, applicable within the range for which the curve of ρ/T against H is straight.

There was until recently some doubt as to whether the Lorenz function was independent of direction in an anisotropic single crystal, but careful measurements by Cinnamon[53], over the whole range of orientations in zinc single crystals, seem to show that it is. In his work, it was verified that the variation of thermal conductivity with angle is in accordance with Voigt's symmetry relations, so that the conclusion drawn from the results may be accepted with some confidence.

Turning to non-metals, a most interesting investigation is that of Crary[54], on

the thermal conductivity of Acheson graphite. This author finds that the conductivity obeys the law $K = B + A/T$, where T is absolute temperature. It would have been very interesting if he had determined the electrical conductivity of the same specimen, in order to find out how the Lorenz function behaves for such a material.

Silica aerogel, a solid insulator with a thermal conductivity less than that of air, has been described[55] by Kistler and Caldwell. It is obtained by removing the water from the hydrogel, and consists of particles with intervening spaces of the order of 10^{-6} cm., i.e. less than the normal mean free path of molecules of air. Since the conductivity of a gas decreases with the mean free path, it is suggested that this accounts for the extraordinarily low conductivity—$5{\cdot}10 \times 10^{-5}$ c.g.s. unit at about 35° C.—and for the fact that the conductivity is not increased by grinding the gel. Evidently, ordinary mechanical treatment would not subdivide spaces as small as these, so that even finely ground powder retains the properties of the material in bulk. By measuring the conductivity *in vacuo*, even lower values ($1{\cdot}47 \times 10^{-5}$ c.g.s. unit) were obtained. The material has a density of the order 0·10 gm. per c.c., and can stand loads of 20 or 30 lb. per sq. in.

A quantity analogous to thermal conductivity—the thermal diffusivity or thermometric conductivity—may be mentioned here. Van Rest[56] has described a method for measuring this quantity in the case of poor conductors, the theory of the method being a slight modification of that applicable to the heat-flow in a semi-infinite solid when the face is suddenly raised in temperature by a fixed amount. In Van Rest's theory, the face of the sample is assumed to vary in temperature according to the law which would apply at a depth x_0 (chosen to suit the experimental results) in the simpler case. This must be a very close approximation to the true state of affairs, and the method should be capable of giving very good results.

§ 4. RADIATION

The various instruments used for radiation measurements—thermopiles, bolometers, radiometers and the like—have been subjected to particularly intensive study during the last few years, but the conclusions reached are not easy to summarise. Much of the work has resulted in increased sensitivity, to such an extent indeed, that Brownian movements in the galvanometer exert the final control on the precision reached. A detailed critical study is to be found in the papers by Cartwright[62], Hardy[63] and others cited in the references[75] given later. Cartwright considers that the most sensitive of all is the vacuum thermopile and that it can be so designed as to reach its theoretical sensitivity.

Guild[64] has described a thermopile for measurements of total radiation in which many sources of unsteadiness and of error are eliminated by connecting two similar instruments differentially, and shielding one of them whilst the other is exposed, an idea used independently but less completely by Beckett.

Bosworth[65] has prepared strips of nickel of the phenomenally small thickness of 0·08 μ, and has applied them to the construction of a very sensitive bolometer.

The method used is that of electro deposition on stainless steel, the nickel film (reinforced by shellac) being peeled off and the shellac then dissolved away with alcohol.

Another method of obtaining high sensitivity has been tried by Harris[66], who uses thermocouples of very low heat capacity, and intercepts the incoming light by means of a rotating sector. The e.m.f. to be measured is thus made intermittent, so that a.-c. amplifiers can be used to magnify it, thus providing the increased sensitivity required.

As regards the use of these instruments in taking observations, Koch[67] has outlined a procedure for the measurement of emissivities, which obviates many of the errors which may creep in if special precautions are not taken, whilst Ritzow[68] has given very full data on the infra-red radiation from a number of oxides, Hoffmann and Willenberg[69] devoting their attention to the ultra-violet.

One of the commonest methods of realizing a "black-body" experimentally is to use a cylinder of length great compared with its diameter. The length necessary to reach a given degree of blackness can only be calculated approximately, and even then as the result of a complicated process. An approximation was given some time ago by Buckley, and was carried a stage further by Yamauti in 1933. Buckley[70] has confirmed many of his results, though they do not agree fully with those quoted by Dorgelo. Accepting them, the ratio of length to diameter for 99 per cent blackness is as follows:

	Emissivity		
	0·75	0·50	0·25
1st approximation	3·8	5·8	9·0
2nd approximation	4·8	7·0	10·7

The question has also received experimental investigation at the hands of Cunnold and Milford[71], whose results are not, however, directly comparable with those of the theory.

On the theoretical side, Finkelnberg[72] has considered the conditions under which black-body radiation can be emitted by a gas, and reaches the rather obvious conclusion that not only is thermodynamic equilibrium an essential, but also that the mass of gas must be so thick that it is a perfect absorber. He applies the criterion to show that the background in the spark discharge cannot be pure temperature radiation.

The experiments of Nichols and Hull on the pressure of radiation have become classical, and it is interesting to learn, after the lapse of 30 years, that some errors of calculation have been discovered[73] in the work. The agreement between theory and experiment is not so good as had long been understood. We may, however, draw double consolation not only from the thought that even Homer has nodded once again, but that in so doing he has saved us from the hypotheses which would have been put forward to bring theory into line with experiment, if the disagreement had been known earlier.

§ 5. CONVECTION AND EVAPORATION

A survey of the progress achieved in the study of convection is difficult to make. The nature of the results, as may easily be realized, is best expressed in a curve or series of curves connecting one or several dimensionless variables, but the precise locations and forms of the curves for bodies of different shapes are hardly of general interest, valuable as they may be when actual calculations are to be made.

As regards processes of measurement, Saunders[76] has described equipment for measuring convection at very high pressures. From arguments of dimensional similarity it is easy to show that such measurements enable us to predict the convective loss at ordinary pressure from bodies of the same shape as those on which the experiments are carried out, but of enormously greater size. This experimental procedure has the advantage, not only of enabling the dimensions of the test specimen to be reduced, but also of making radiation negligible in comparison with the convective heat loss. Its chief disadvantage is that the thermal conductivity of the gas at the high pressures enters the equations of reduction, and present information on this quantity is rather scanty.

A number of authors, among whom we may mention Ramdas and Malurkar[77], and Hencky and Neubert[78], have adopted optical methods from which to deduce the heat transfer by convection. In most cases they in effect measure the refractive index of the gas by interferometer methods, though Fishenden and Saunders adopt the simpler, and apparently very effective, plan of deducing the refractive index from the deviation of a beam of light. Among those who have used interferometry with conspicuous success may be mentioned Kennard[79], whose paper contains some excellent photographs of fringes produced in heated air, and Hansen[80], who also shows photographs.

Another application of optical methods makes use of the difference between the refractive indices of cool and of heated air, in order to make the actual convection currents visible. It thus gives no quantitative information either on the velocities or the heat losses. This method known as the Schlieren method (on account of its original use in examining blocks of optical glass for flaws), has been brought to a high degree of perfection by Schmid[81] in Germany and by Townend[82] in England. The former author shows several pages of photographs of the air flow around various heated bodies, which are remarkable for their clarity of detail.

In the general correlation of results, a variety of dimensionless variables are in use. They are commonly, especially in Germany, referred to by the first two letters of the name of a worker in the subject. Thus Reynold's number is $Re = Vd/\nu$, Péclet's is $Pe = Vd/\alpha$, Grashof's, $Gr = gd^3\beta\theta/\nu^2$, Prandtl's is $Pr = \nu/\alpha$, and that of Nusselt is $Nu = \alpha d/k$. In these formulae, V is the velocity of the stream of fluid, d a length parameter, ν and α are the kinematic viscosity and thermal diffusivity of the fluid, g is the acceleration due to gravity, β and k the expansion coefficient and thermal conductivity of the fluid, whilst θ is the temperature excess. The manner in which these variables enter into problems of free and forced convection differs considerably. A typical example of the way in which they succeed in corre-

lating the results for the latter condition may be found in the experiments of Griffiths and the writer[83] on the heat transfer to air from metal pipes arranged in so-called "batteries". These experiments also illustrate the great increase in heat transfer which accrues when the motion of the fluid changes from laminar to turbulent flow. For free convection, the discussion given by King[84] is most striking, in that a single curve is found to cover the enormous range of Grashof's number from unity to one hundred million. It may also be of interest to mention the careful work of Jodbauer[85] on the temperature distribution around the circumference of a pipe from which heat is being lost by forced convection.

In cases where fluid flows turbulently parallel to a solid wall, various workers have noted a difference in the heat transfer coefficient according as the fluid is hotter or cooler than the wall. Kaye and Furnas[86] have successfully explained this as due to the lower viscosity of the boundary layer of liquid, which is therefore much thinner, in the cases where the liquid is hotter. Their explanation is quantitative and correlates the results of a number of workers. Moreover, it explains the curious fact, observed by Furnas, that for gaseous fluids (for which viscosity increases with temperature) the discrepancy is of opposite sign to that in liquids.

The only theoretical treatment of value known to the writer during the period under review is that of Piercy and Schmidt[87] who deal particularly with the case of fluid flowing over a flat plate. They adopt as basis, the approximation to the equations of motion, due to Oseen. This takes account of viscosity, and in part also of inertial effects. It would be expected from its nature that the approximation would be valid only for small Reynolds' numbers, but in fact Piercy and Schmidt find that the results of their analysis agree quite well with experiments, such as those due to Fage and Falkner, at moderately high values of this variable.

The process of heat transfer when a liquid in a tube is evaporated by heat applied outside, is only just beginning to receive really scientific study. A surprising fact stated by King[84] (p. 194) is that in the case where a refrigerant is being evaporated, "there is no discontinuity at the metal-liquid interface, but the liquid in the surface film is heated considerably above the boiling-point. The greater part of the temperature drop from the metal to the main body of the liquid occurs in this superheated film". An almost exactly similar result was found for steam by Jakob and Fritz[88]. They state that when water evaporates there is a temperature difference between the liquid and the vapour, but from pressure measurements it can be shown that the latter is at the true boiling-point. Consequently, the liquid is at a temperature above this, and in fact it is demonstrable that the superheat is mainly in the surface layer. A very clear account of some recent work on heat transfer through flat plates to evaporating water is that given by Jakob and Linke[89] in 1933. In the converse process of condensation, which was discussed in last year's report, it may be of interest to add that the heat transfer when superheated steam condenses, is approximately equal to that obtained with saturated steam at the same pressure (and therefore at a lower temperature). In condensation, two regimes are possible: the formation of a skin of liquid, or the production of droplets. The factors affecting which mode will occur in any particular case were

discussed by Jakob[90] in 1932. The process was subjected to theoretical treatment a few years ago, by Nusselt and others[91], and Jakob finds that in general his results are in harmony with theory. Hebbard and Badger[92], on the other hand, find that whilst the form of the curve (for the case of steam condensing on the outside of a vertical pipe) is similar to the theoretical one, yet the results are higher by about 33 per cent. Indeed, as Kirkbride[93] has pointed out, the major portion of data on condensing vapours on vertical tubes, are higher than predicted by Nusselt's theory. Kirkbride is of opinion that Nusselt's theory does not hold for values of Reynolds' number greater than about 2500, and he is not inclined to accept the explanation favoured by some writers, that the increase in heat transfer which then occurs is due to the onset of the regime of drop-formation.

A somewhat different aspect of the problem of evaporation is that studied by Houghton[94], who measured the rate of decrease in diameter of water drops in still air. He finds that for a range of temperatures and of humidities, the equation $\rho dv/dt = -2\pi ka(\sigma_s - \sigma_\infty)$ holds. Here ρ is the density of the liquid, and σ that of the vapour, the suffixes s and ∞ referring to the neighbourhood of the surface and to a great distance from the drop respectively; v is the volume of the drop, a its radius and t the time. He finds that the constant k is nearly equal to the coefficient of diffusion of water vapour in air. The investigation has obvious relations to the theory of the wet-and-dry-bulb hygrometer, and in fact the equation which Houghton finds is in agreement with the one usually assumed in the theory of that instrument. It is to be noticed that in calculating the vapour density, the temperature of the drop is to be taken as lower than that of its surroundings, owing to the cooling effect of the evaporation itself. This effect, together with a number of other causes of departure from the simple theory, was mathematically treated by N. Fuchs[95].

§ 6. VAPOUR PRESSURE

During 1934 a new formula and set of tables for the vapour pressure of water[101], based on previously published data, has been put forward from the American National Bureau of Standards. It differs but little from other tables, though it is desirable in the interests of uniformity that those who have occasion to use values of high precision should utilize it, unless there is some special reason to the contrary.

The accuracy of the data on which such an equation is based, is well illustrated in the following table, which is condensed from one given by Jakob. It is to be remembered that this table includes results both by the static and the dynamic methods.

Another fluid to which attention has again been devoted is mercury. Eck has given an enthalpy-entropy diagram[102] and an enthalpy-pressure (Mollier) diagram for this element, covering the pressure range from 0·001 to 100 kg./cm.^{2}, i.e. temperatures up to 700° C.

The vapour pressure of calcium was measured by Rudberg[103], using an effusion method developed from that first used by Knudsen many years ago. His results between 500 and 625° C. agree with an equation $\log_{10} p = 8{\cdot}15 - 9670/T$

Vapour pressure over water

Temperature (° C.)	Pressure (kg./cm²)			
	Keyes and Smith, 1901	Smith and Keyes, 1933	Egerton and Callendar, 1932	Osborne and collaborators, 1933
100	1·0332	—	—	1·0332
130	2·7548	—	—	2·7544
160	6·3030	6·3034	—	6·3027
190	12·796	12·802	12·812	12·800
220	23·641	23·659	23·664	23·658
250	40·523	40·557	40·556	40·560
280	65·419	65·452	65·472	65·458
310	100·66	100·62	100·65	100·64
340	149·12	148·92	148·94	148·96
370	214·78	214·64	214·74	214·68

where p is in mm. of mercury and T is absolute temperature, but they disagree with the earlier measurements of Pilling by a factor of about ten. Pilling used a method in which the rate of evaporation was determined by weighing, and thus the theoretical basis of his procedure is more doubtful than that of Rudberg. That the values obtained by the latter are at least approximately correct seems to be shown by the fact that they give the correct value of the "chemical constant" in the vapour pressure equation. On the other hand, the data do not extrapolate well to join the curve obtained by Hartmann and Schneider at higher temperatures. These observers used a direct dynamic method, which should not be open to serious doubt. It thus appears that confirmatory work would be desirable.

In connexion with vapour pressures, a curious paper by Trautz and Ader[104] was published during the year, and was the cause of two further papers by other authors[105, 106], who discuss it from different points of view. The argument in the first paper, that of Trautz and Ader, is that the vapour pressure curve is such a useful concept in many thermodynamic calculations that it would be useful to find some analogue to it in the region above the critical point. Now in any region, the equation of state gives the relation between p, v and T. Along the saturation curve, this reduces to a relation between p and T only, so that the analogue sought is also a relation between p and T, i.e. it is a curve of constant volume. The only such curve meeting the vapour-pressure curve is that for which the fixed volume is the critical volume, so that if an analogue exists, this curve and no other must be it. Trautz and Ader then proceed to plot the p-T curve for which the specific volume is equal to the critical volume, in the case of CO_2, and they find that it is practically linear, and is tangential to the vapour-pressure curve. This evidence is based on their own observations of the p-T relation at the critical volume as determined experimentally by them, and is taken to prove that this p-T curve is a smooth continuation of the vapour-pressure curve; they do not, however, define the meaning of this phrase, i.e. to what order of derivatives the two curves are supposed to agree. Also, they do not test the agreement between their experimental results and those deducible from standard tables of the properties of CO_2.

Verschaffelt's criticism of this[106] is that the continuation of the vapour-pressure curve is calculable from the equation of state, as the locus of points at which the pressures, temperatures and Gibbs' functions for the two phases are equal. This locus is real above the critical point, although the volumes are imaginary. Verschaffelt also remarks that for certain equations of state the curve so defined is identical with that adopted by Trautz and Ader as the continuation of the vapour-pressure curve.

Eucken[105], on the other hand, considers that the natural continuation of the vapour-pressure curve is the locus of maxima of $(\partial I/\partial p)_T$ where I is the enthalpy, i.e. it is the locus of maxima of the isothermal throttling coefficient. He supports this with the experimental fact that the measured maxima for methane do in fact lie on a "continuation" of the vapour-pressure curve. Like Trautz and Ader, he fails to define with precision what is meant by a "continuation". Eucken also gives a discussion of the matter from the standpoint of the equation of state, from which he concludes that there are several continuations, e.g. the maxima of $(\partial v/\partial T)_p$, $(\partial v/\partial p)_T$, etc. He also discusses the experimental steepness of the maxima and minima, above the critical point, of those quantities which have discontinuities below that point. In this discussion, he comes very near to the views of Callendar[107] as to the phenomena immediately above the critical point, though he does not actually either adopt or refer to them.

§ 7. LIQUIDS, VAPOURS AND GASES

The properties of helium in the liquid form have been of interest ever since it was first discovered that two forms exist, between which there is a transition point at the low temperature of 2·19° K. (or rather, a transition curve in this neighbourhood). One interesting feature is that He II shows a negative coefficient of thermal expansion[110,111]. This phenomenon was observed by Kammerlingh Onnes and Boks some years ago; they thought that there was also a discontinuity in the density itself at the transition point, but it has since been shown by W. H. Keesom and Miss A. P. Keesom[110] that this is improbable.

That a definite energy change occurs in the transition is proved by the measurements of specific heat[112] by W. H. Keesom and Miss A. P. Keesom. These authors find that the specific heat at saturation pressure rises from about 0·2 to 3·0 cal. per gm. per 1° C., and then falls again. In its reduction towards the final value of 0·54, it has fallen as far as 1·1, certainly within 0·02° K., and probably within a few thousandths of a degree, of the temperature at which it shows its maximum. The thermodynamic relation which connects dL/dT, where L is latent heat of evaporation, with the specific heats, then predicts a "hump" in the latent heat curve, which agrees exactly with what was observed experimentally.

Although we have spoken of an energy change, it is not to be understood that there is a discontinuity in the energy as helium passes from one form to the other. All that is proved by the observations on specific heats is that the gradient of the energy-temperature curve has a discontinuity. Actually, the evidence is to the

effect that neither specific volume nor energy has a discontinuity. Hence the first derivatives (but not the second) of the thermodynamic potential are continuous. Halpern[113] has claimed that in such a situation it is impossible for the two phases to coexist in equilibrium. To illustrate this, he imagines a melting-point curve plotted in an energy-volume diagram. With an ordinary case of melting, where the phases differ in energy-content, a slight variation in temperature alters the amount of each phase, and hence the volume and energy, in such a way that the representative point moves a small distance along the curve of separation. In the case of He I and He II, however, an infinitesimal variation does not alter the volume or energy, so that the point cannot adjust itself by travelling along the curve; it must move off it into the region either of He I or of He II. In other words, the mixture of two phases cannot be in stable equilibrium. This agrees with what is observed experimentally. On the other hand, the mathematical details of Halpern's argument have been criticized by Guggenheim[114], on what seem to the present writer to be good grounds.

Keesom points out that the excess energy as between He I and He II might be due either to a change within the atom itself, or in the motions of the atoms, or in their mutual arrangement. He considers that the equality of the atomic polar moment for the two forms of the liquid rules out the first possibility, whilst the fact that we are dealing with an atom in which rotation and internal vibrations cannot occur, renders the second improbable. We are thus left with the probability that the change is one in the mutual arrangement of the atoms. He points out that the transition occurs at a reduced temperature which lies between those at which solidification occurs in such gases as oxygen, hydrogen and neon, and suggests therefore that it is in the nature of an analogue to solidification. Although not specifically stated, this suggestion is equivalent to a suggestion, either that the cybotactic state, or else that of a liquid crystal, is assumed at this temperature.

The cybotactic theory of liquids[115] just mentioned, which asserts that in them the molecules are arranged with some degree of regularity, and not in the random manner assumed in early theory, dates from about 1929. The first observation bearing on the point was that of Debye and Scherrer in 1916, when they found that diffraction haloes could be obtained by passing X rays through a liquid. Now, since the ultimate scattering particles are the electrons, the experiments indicate either a periodicity within the atom, or within the molecule or finally one characteristic of the spatial groupings of the molecules themselves. The experimental evidence seems to show that the last effect is the one which is actually measured in practice, and it follows that in a liquid there must be definite ordered groups of molecules. It does not follow that *all* the molecules are ordered in this way, nor that the groups are permanently composed of the same individual molecules.

In experiments by Ward[116], consisting of X-ray examination of mixtures of liquid benzene and cyclohexane, the intensity of the rays diffracted at angles up to 9° was measured. For cyclohexane there is a peak at 4·00°, the intensity here being about seven times that of the background.

Benzene shows a peak at 4·33°, of intensity about six times that of the back-

ground. These observations alone show that there is some regular structure in the liquids, and calculation in fact shows that the preferred spacing is the effective thickness of the benzene and cyclohexane rings respectively as found in crystals. Thus the spacing is probably intermolecular and not intramolecular.

Ward examined also solutions containing from 20 to 70 per cent cyclohexane in benzene, and found that in these the peaks characteristic of both components are clearly distinguishable (though not actually resolved in his apparatus). Thus he pictures the mixture as a sort of ultra-colloid, the cybotactic groups of both materials being present, though they are so small that the Tyndall effect (scattering of light) could not be observed.

Katzoff[117] has pointed out that the groups postulated by Ward need only be aggregates of about 10 molecules each, so that the conception of "association" is quite as relevant as that of colloidal "emulsion". Katzoff's own measurements were made over a much wider angular range, and, from the more complete intensity curves which he obtained, he finds that for water, for example, each molecule has, on the average, about four nearest neighbours. This is interesting, in view of the fact that in ice each oxygen atom is tetrahedrally surrounded by four others. For heptane and decane, he interprets his results as showing that, except for the parallel orientation of the molecules, there is extensive disarrangement in the liquid. In summarizing his conclusions (based on the detailed study of five different liquids) he says that "the view point which has been found useful—is that, to a predominating extent, adjacent molecules in the liquid are held together in very nearly the same manner as in the crystal, while, except for this, the arrangement is random".

It should be mentioned that Parthasarathy[118], as also Gringrich and Warren[119], adopt a somewhat different interpretation of the X-ray diffraction effects, but that, on quite different grounds, Andrade has also found it desirable to introduce the conception of a partially ordered system in liquids.

The X-ray diffraction method throws light also on the processes which occur in a fluid in the neighbourhood of the critical point. This subject has been investigated by Spangler[120] for ethyl ether, and by Benz and Stewart[121] for isopentane. Both agree that the form of the intensity curve is fixed only by the specific volume of the fluid. Thus the degree of orderliness is determined by the average closeness of packing of the molecules, and is independent of their thermal motions. This is true even when the comparison is made between liquid and vapour, at pressures and temperatures such that both have the same specific volume. An interesting difference between the two fluids occurs, however, in the specific volume at which the curves change from the type characteristic of a liquid to that characteristic of a gas. For ethyl ether, this specific volume is about equal to the van der Waals' critical volume, whilst for isopentane, it is nearly twice the critical volume. It thus appears, unexpectedly, that the non-polar liquid forms more stable groups than the polar one.

In connexion with the theory of Callendar and others as to the nature of the phenomena at the critical point[107], some highly accurate determinations of the density of propylene vapour and liquid[122], under saturation pressure, should be mentioned. The observations cover ranges quite close to the critical point, and there

is no indication that the curve has other than the blunt shape usually assumed, with a vertical tangent at the critical point. Koch[123] has also measured the total heat of steam above the critical point (374° C.) and fails to find the latent heat of evaporation which would be required on Callendar's theory. He also fails to confirm Callendar's statement that traces of air influence the heat content values. It should be mentioned, however, that Koch, like earlier workers, finds a marked maximum in the specific heat at about 380° C., and that his maximum is higher than that of earlier workers. This certainly suggests that, with still greater care in removing dissolved air, a latent heat might be found.

Another aspect from which we must consider work on gases is that of the determination of the characteristic equation for real gases. It is well known that the behaviour of a gas, as its density tends towards the zero limit, represents the behaviour that it would exhibit if it remained "perfect". In particular, the molecular weight can be determined with an accuracy quite as high as that obtaining in chemical or in mass-spectrographic work, provided that the extrapolation can be carried out sufficiently exactly. The subject of this extrapolation has been considered by Birge and Jenkins[124] during the year under review. They point out that, basically, the problem is that of finding the limiting value of $L_p = w/pv$ as $p \to 0$, where v is the volume of a mass w of gas at pressure p. Now when p is quite small, the variations of L_p are such that if it is (experimentally) a linear function of p, then $1/L_p$ is also linear to within observational accuracy. The two extrapolations, however, may give slightly different results, but the difference between them cannot by its nature be of significance; neither result is known to more figures than the first at which they begin to disagree.

Interesting examples of the use of these methods for molecular weight determinations are to be found in the determinations by Heuse and Otto[125] of the atomic weights of krypton and xenon. For the former gas, they find $83{\cdot}66 \pm 0{\cdot}01$, which may be compared with the mass-spectrographic value 83·77, and for the latter they find $131{\cdot}22 \pm 0{\cdot}02$, whereas Coppock and Whytlaw-Gray[126] find $131{\cdot}26 \pm 0{\cdot}005$ by a similar method, the mass-spectrographic value being 131·27.

Another example is to be found in the work of Stewart and Maass[127], who, by means of accurate density measurements, have measured the molecular weight of sulphur dioxide. They find that the curve of apparent molecular weight against pressure is not straight, so that the extrapolation is one which requires higher order terms. Their result for the molecular weight is 64·075.

§ 8. HYGROMETRY

Williams[134] has designed a convenient form of dew-point instrument of the type in which the dew-point temperature is measured by means of a thermocouple attached directly to the metal surface on which dew appears, thus avoiding any error due to temperature discontinuity between the cooling liquid and the metal. He also avoids temperature lag by using a fine spray of ether to cool the surface, instead of the more usual arrangement where several c.c. of liquid are present.

More ambitious dew-point instruments are those which operate by means of a photoelectric cell. An excellent example of this type is that described by Moss[135], in which diffraction of a beam of light by the droplets of dew is utilized to make the apparition of the dew more easily visible. The instrument which he describes can be used with visual observation, but the use of a photoelectric cell offers the advantages, not only of making recording easy, but also of offering a means of automatic control of humidity.

The wet-and-dry-bulb hygrometer has been studied experimentally[136] by Griffiths and the writer, tables of the relative humidity corresponding to given wet and dry bulb readings up to 100° C. having been obtained for the first time. The theory of this form of hygrometer depends closely on that of evaporation from moistened surfaces, and a discussion of the results found in these experiments, in the light of modern theory, was given by Whipple[137].

In connexion with hygrometry, we may note also the new determination of the vapour pressure of water over solutions of sulphuric acid and other substances at various temperatures and for various concentrations[108, 109]. These data, apart from their use in physical chemistry, are of value when a stream of air of known and constant humidity is desired, since one effective method of attaining this end is to bubble air through such solutions.

Most hygrometers other than the dew-point instrument measure relative humidity rather than vapour pressure. In cases where the latter is the quantity desired, Wenk[138] has shown how by quite simple link mechanism, an instrument reading vapour pressure directly can be constructed with a bimetallic strip and a hair hygrometer. Since H, the relative humidity, is p/p_0 where p is the actual and p_0 the saturation pressure, we have $p = p_0 H$ or $\log p = \log p_0 + \log H$. Now p_0 varies with the temperature T in such a way that $\log p_0$ is approximately $\alpha + \beta T/(\gamma + T)$ or on expansion, $\log p_0 \simeq \beta\,(T/\gamma - T^2/\gamma^2) + \alpha$. Consequently, the bimetallic strip is made to give a reading distorted in the manner determined by the quadratic expression on the right, whilst the hair element also moves the pointer independently, the actual reading being the sum of those due to the two effects. Now, granting that the extension of the hair is proportional to the logarithm of the relative humidity, it follows that the total movement of the pointer is proportional to the logarithm of the vapour pressure, i.e. it gives the vapour pressure on a logarithmic scale. The instrument can easily be used as a recorder, marking on a chart.

No investigator has yet reduced the hair hygrometer to rule completely, although qualitatively its behaviour at extreme temperatures and extreme humidities can be predicted from the work of Griffiths and others. A further attack on the problem has been made by Grundmann[139], who studied *inter alia* the effect of temperature changes from −60 to +60° C., at each of a number of humidities. He finds for example at 40 per cent relative humidity that the extension of human hair alters from 65·3 to 59·9 as the temperature falls from +60° to −60° C., but that the effect of temperature is much less at high humidities. Grundmann has also investigated the effect of other gases, particularly acids, in the atmosphere, and has

considered the properties of the hair of different animals. He is of opinion that human hair is the most reliable for the purposes of hygrometry. Schweitzer[140], on the other hand, has made a most exhaustive study of the properties of cellophane, which can be adapted to hygrometry in various ways.

§ 9. THERMODYNAMICS, INCLUDING CHEMICAL REACTIONS AND THEORY OF SOLUTIONS

The nature of the phenomena which occur when gases react in a closed vessel has long been a source of controversy, and cannot be entirely omitted from a report like the present, although it is difficult for one who has not worked on the subject to give a satisfactory review. Many workers, beginning with Pier in 1908, have used measurements of the maximum pressure developed, for the purpose of calculating specific heats, but others have claimed that the results are incorrect. The second school holds that combustion at any point is incomplete at the time that the maximum pressure is reached there, the phenomenon (if it really exists) being referred to as "after burning". The main evidence for this theory has been the fact that David and others found the ratio of pressure increase (after correction for heat losses) to be greater in a large vessel than in a small one. Lewis and von Elbe[143] have repeated these experiments with particular attention to certain details which they consider may have been the origin of the results found by David, and they now find almost perfect agreement for this ratio, using vessels of different sizes. They also discuss other, less direct evidence, which has from time to time been put forward in favour of the theory of "after burning", and apparently show that the arguments were invalid.

A like conclusion was reached by Maxwell and Wheeler[144], who repeated some experiments of David and El Din, but with an explosion vessel intermediate in size to the two which they had used. They found pressure rises agreeing well with those in the larger vessel used by David and El Din, whilst they point out also that the results of Thorpe for a vessel 45 cm. in diameter differ from those obtained by David and El Din in a vessel of the same size by almost as large an amount as the difference which these latter investigators found between their large and small vessels.

As regards pure thermodynamics, the precise meaning of the third law is becoming clearer with time.

The dispute arose originally, not as to whether the law is ever true, but whether it is ever untrue, and if so, under what precise circumstances. Thus Nernst, Planck, Einstein, and indeed all the pioneers of the subject, accepted without question that the entropy of pure crystalline substances vanishes at the absolute zero. The work of Giauque and others about 1920 seemed to show that the law did not apply to "vitreous" solids (undercooled liquids); most recently it has been found that even such an obviously "pure crystal" as solid hydrogen has a zero-point entropy, and it became doubtful whether the law ever applied at all. The position appears to be best elucidated by an appeal to statistical mechanics. The entropy is generally taken

to be $k \log Z$ where Z is the number of distinguishable configurations; let us modify this to the extent of allowing Z to be the number of possible configurations which might be present *as far as our knowledge goes*. Thus in hydrogen (a mixture of para- and nine kinds of ortho-hydrogen), each kind of ortho-hydrogen might be present with molar fraction from 0 to $\frac{1}{4}$. Thus Z is $9^{3N/4}N!/(N/4)!\,(3N/4)!$ This gives a zero-point entropy agreeing with Giauque's experimental result.

From this point of view, a solid solution should have a zero-point entropy of $R(X_1 \log X_1 + X_2 \log X_2)$ where R is the gas constant and X_1, X_2 are the molar fractions in which the two constituents are present. It is interesting to note that in a very careful research Eastman and Milner[145] found experimentally the value 1·12 for the zero-point entropy of a solid solution containing 72·8 per cent of silver bromide to 27·2 per cent of silver chloride, where the above formula gives 1·16.

Their work is sufficiently interesting to be worth detailed quotation. They first showed that the crystals were identical whether made by displacing some of the bromine from silver bromide by chlorine, or conversely; also that the refractive index lay between those of the pure substances, with no sign of inhomogeneity. It was also checked that prolonged exposure to liquid hydrogen temperatures did not lead to any separation of the constituents, and that the X-ray diagrams indicated a true solid solution, the lines of each of the original salts being entirely absent. The change in thermodynamic potential on formation was determined by e.m.f. measurements to be 254 cal. per mol. The heat of formation was determined calorimetrically to be 80·7 cal. per mol. at 25° C., whilst the specific heats of the pure salts and of the solid solution were determined from 15° K. to 298° K., extrapolation to 0° K. being made by means of Einstein and Debye functions.

From the figures, the entropy of mixing at 298° K. is 1·12 cal. per degree.

Now the specific heat of the solution does not vary from that for the appropriate quantities of the constituents, as may be seen from the following table given by the authors.

log T...	1·25	1·50	1·75	2·00	2·25
C_p: Solid solution	2·000	4·945	8·295	10·660	11·920
Constituents	1·982	4·899	8·240	10·679	11·921
Difference	0·018	0·046	0·055	−0·019	−0·001

It is extremely unlikely that any variation from such a simple law would set in at still lower temperatures, so that it is almost certainly established that the solid solution has an entropy of $1 \cdot 1_2$ units at the absolute zero. Eastman and Miller link this with the ideas set out above, by pointing out that every interchange of a chlorine and a bromine atom gives a new configuration. Either salt alone, on the other hand, would have zero entropy at absolute zero, since no such interchanges are possible.

As to the general question, these authors consider that all the anomalous cases (solid solutions, mixtures where isotopic atoms provide the variety of configuration, variations in nuclear spin, etc.) correspond to the existence of a continuous (even

if narrow) band of energies, and that consequently the law is accurately stated if restricted to "any phase of sharply specifiable energy".

This criterion is analogous to one which has been put forward by several authors, who propose to restrict the statement of the third law to phases which are in their lowest energy states.

The bearing of the uncertainty principle on the question has been discussed by Rhodebush[146], who reaches the conclusion that the entropy of any phase vanishes at absolute zero, though there may be difficulty in defining precisely the entropy at sufficiently low temperatures, for those substances which do not reach their lowest energy states at attainable temperatures.

As regards the Debye-Hückel theory of electrolytes, there are no significant advances to chronicle, but it may be noted that the basis of that part of the theory which has received criticism in the past, has been reviewed and put on a firm statistical-mechanical basis by Kirkwood[147].

Some light is thrown on the peculiar phenomena connected with the absorption of gases on solid surfaces by the conception of "active molecules". Thus Beebe and Wildner[148] find by measurement of the heat of absorption, that when carbon monoxide is admitted in small quantities into contact with clean copper, the energy changes are much greater for the first portions than for the later ones. This indicates clearly that there are certain particularly active centres present, which later become exhausted.

A similar conclusion is deducible from the experiments of Zickermann[149], who measured the rate of fall of pressure on admitting gas first into an empty vessel, and then into one containing glass or mica. The difference in rate gives the absorption on the glass or mica. His experiments were carried out at very low pressure (of the order of 10^{-4} mm. of mercury) and at low temperature, the gases used being nitrogen and argon. He finds that the rate falls off, indicating the presence of active centres, which he estimates to be present to the extent of about 1 per cent of a monomolecular layer. Since he obtained data at three different temperatures, the usual thermodynamic equation can be applied to calculate the heat of absorption, which comes out as follows.

Heats of absorption of gases on solid surfaces (cal. per mol.)

	Nitrogen	Argon
Glass	337	125
Mica	1830	571

The thermodynamics of membrane equilibria has been further studied by Donnan[150], and also by Gatty[151], who has been concerned particularly with the bearing of the phase rule on the question.

In thermo-chemistry, there is a most valuable paper by Becker and Roth[44], which contains data on the heat evolution in some ten or twelve reactions, the ultimate object of the work being to obtain the heats of formation of ammonia and

of nitric acid. The more interesting of the individual results, using square brackets to denote solids and parentheses for gases, are:

$$[NH_4NO_3] + Aq = NH_4NO_3, Aq - (6{\cdot}32_6 \pm 0{\cdot}01_0)$$
$$HNO_3, Aq + (NH_3) = NH_4NO_3, Aq + (20{\cdot}89_5 \pm 0{\cdot}01_1)$$

From these and other reactions, Becker and Roth calculate that

$$(N_2) + 2\tfrac{1}{2}(O_2) + H_2O + Aq = 2HNO_3, Aq + 14{\cdot}84$$

and

$$\tfrac{1}{2}(N_2) + 1\tfrac{1}{2}(H_2) = (NH_3) + (11{\cdot}01 \pm 0{\cdot}07).$$

All the figures are in k.-cal. per mol., and refer to constant pressure and to 20° C., the nitric acid being $N/20$. Thomsen's result for the formation of nitric acid was 14·9, whilst Berthelot obtained 14·3, and the Bureau of Standards recently gave 14·55 for a concentration $N/4$. The heat of formation of ammonia gas was determined by Berthelot and by Thomsen, but their results are worth little. Of greater interest for comparison are the results obtained for this heat of formation by Haber (11·08) and by Kraus and Ridderhof[36] (11·64). The latter investigators worked at −34° C., and the result given is corrected to 20° C. by Becker and Roth.

§ 10. SPECIFIC HEATS OF GASES

In the past, the explosion method has frequently been used to determine the heat capacity of the gases resulting from reactions conducted explosively. Lewis and von Elbe[155] have now shown how the same data may be used to determine the heat capacity of the reactants. For this purpose, they analyse the phenomena occurring within the explosion vessel when this is of spherical form and the reaction starts at the centre. In this case, the reaction naturally proceeds outwards as a spherical wave, so that at any instant the unburnt gases consist of an outer shell, subjected in general to adiabatic compression, and consequently raised in temperature. A knowledge of this rise, together with that of the work done in compressing the gas (which follows from a knowledge of the pressure and temperature in the inner zone), gives the specific heat. At first sight the method would seem to be one of great experimental difficulty, and to suffer from relying too much on a detailed theory of some very complex processes, but the authors mentioned seem to have proved that these fears are unfounded, provided at any rate that the temperatures can be measured with the necessary accuracy.

The theoretical synthesis of energy values from molecular constants, as revealed in the first place by investigations on infra-red structure, has been actively continued. Thus Johnston and Long[156] give tables of the specific heat from 0 to 300° K. of hydrogen molecules containing heavy hydrogen, and the ortho-para equilibrium for D_2. These figures were subsequently confirmed indirectly by the experiments of Farkas and Farkas[157]. Johnston and Long also give the dissociation constants and the entropies of HD, D_2 and H_2 from 1000 to 3000° K. They note that calorimetric determinations of these entropies at very low temperatures are desirable, in order to establish the exact manner in which the third law of thermodynamics is to be applied to such solids. As explained in the paragraph dealing with that

subject, it is generally accepted that there is a zero-point entropy for many substances. This question was discussed by Zeise in a series of papers[158], which contain also a review of work on spectroscopic determinations of specific heats. Clusius[159] has since shown that there is a zero-point entropy in certain cases where Zeise claimed that there was no evidence for it.

The application of similar methods for calculating the specific heats of polyatomic molecules is now being actively developed, and Gordon[160] has given a detailed discussion, with tables of numerical results, for steam. The analysis of the band spectrum of course provides the data necessary for these calculations, so that agreement with thermal measurements provides an independent check. In the present instance, Gordon combines his data with Rossini's measurements of heats of reaction, to deduce the equilibrium constant in the water-gas reaction. The agreement with direct measurements of this constant by Neumann and Walker (1928) and with indirect ones by Emmett and Schultz (1933) is very good. It is interesting to compare Gordon's values for the specific heat of steam with those given by Partington and Shilling as the best in 1922, and with the values given in Callendar's Steam Tables, as quoted in Ewing's *Thermodynamics for Engineers* (1st edition).

Temperature (° C.)	Specific heat (cal./mol.)		
	Gordon	Partington and Shilling	Callendar
25	8·00	—	—
100	8·11	8·72	(9·0)
200	8·32	8·63	8·75
300	9·57	8·62	8·66
400	8·84	8·65	8·63
500	9·13	8·74	—
600	9·42	8·87	—
800	10·02	9·29	—
1000	10·59	9·89	—

Another technically important gas for which values of the specific heats have been calculated (up to 3500° K.) is carbon dioxide. Here also by combining the results of the calculations with measured thermal quantities, the equilibrium constant in the reaction $CO_2 = CO + \frac{1}{2}O_2$ has been deduced and found to agree with direct experiment.

The question of the agreement between these spectroscopic values for the specific heats of gases, and those deduced experimentally, is an important one. At low temperatures, where calorimetric methods may be used, the agreement is satisfactory. At higher temperatures, where the velocity of sound method becomes more suitable, appreciable disagreement had been noted, as mentioned both in the Heat and the Acoustics sections of last year's *Report*. This is commonly attributed to the need for a finite time before the vibrational energy reaches equilibrium with the translational, though there are some who believe that the phenomena are more complicated. However this may be, Sherratt and Griffiths[161] have investigated the

ratio of specific heats of carbon monoxide up to 1800° C., using two different frequencies. They find that the experimental results calculated for either frequency alone differ significantly from the spectroscopic values, but that agreement is obtained if they extrapolate to zero frequency.

The most thorough discussion yet given on this question in the light of experimental results is probably that in a paper by Eucken and Becker[162]. They studied chlorine and carbon dioxide at three frequencies, and over a range of temperatures from −30 to 140° C., examining also the effect of many different impurities. They thus obtained information on the effectiveness of the various molecules in exciting vibrations, as well as data on the relaxation time for the Cl_2 and CO_2 molecules. The latter quantity decreases by a factor of about 3 over the temperature range mentioned. It appears also that some 50,000 collisions must occur in CO_2 at room temperature before a molecule receives a quantum of vibrational energy, but that this number is considerably reduced when highly reactive impurities are present.

Again, the explosion method can be called on for evidence as to the truth of the theoretical values. The question is discussed in detail in a paper by David and Leah[163], who come to the conclusion that the results obtained by this method definitely substantiate the theoretical values.

In connexion with the specific heats of gases, it is surprising to note that the specific heat of argon has never been measured above 25° C. Heuse measured it from −180° C. to room temperature, and found a linear *decrease* of 5 per cent. Since simple theory asserts that it should be constant, and since it is assumed in the explosion method that its value at 2000° C. is the theoretical value, there seems good reason for an immediate experimental investigation at some elevated temperature.

§ 11. DILATATION

The most interesting work in connexion with thermal expansion is almost outside the borders of the subject of "heat". The question which arose was whether the thermal expansion of the lattice of metallic crystals, as determined by measuring the increase of lattice parameter, agreed with the macroscopic expansion.

Owen and Yates[167] by X-ray methods find that in the case of zinc there is close agreement over the range 20 to 100° C. with the metrological measurements of Grüneisen and Goens, and that this agreement even extends to the longitudinal and transverse expansions separately. It is also of interest to note from their measurements that the lattice expansion continues regularly to within 4° C. of the melting-point, the highest point at which they could work.

Similarly, Jesse[168] finds that the lattice expansion of nickel agrees with the macroscopic value from 450° C. right up to 1200° C.

For bismuth, Jay[169] finds excellent agreement between the X-ray measurements and the known macroscopic expansion. His observations extend to within 2° C. of the melting-point. They show a marked decrease in the expansion coefficient at about 245° C., as well as a distinct increase in passing through the transition point at 75° C.

As regards gases, a most careful series of measurements of the volume coefficient for a number of gases has been carried out by Coppock and Whytlaw-Gray[126], who find that the effect of absorption of the gas on the walls of a vessel may affect the measurements quite seriously, particularly in the case of a condensable gas.

REFERENCES

NOTE. *The bibliography is divided into sections, at the end of each of which will be found items to which no reference is made in the text, owing to limitations of space*

§ 1. TEMPERATURE

(1) KÜRTI, N. and SIMON, F. *Nature*, Lond., **133**, 907 (1934).
(2) JACKSON, L. C. *Low Temperature Physics* (1934). London: Methuen.
(3) KEESOM, W. H., VAN DER HORST, H. and JANSEN, A. F. *Commun. phys. Lab. Univ. Leiden*, No. 203*b* (1930).
(4) KEESOM, W. H., BIJL, A. and VAN DER HORST, H. *Proc. Kon. Akad. Wet. Amst.* **34**, 1223 (1931).
(5) HEUSE, W. and OTTO, J. *Ann. Phys.*, Lpz., **9**, 486 (1931).
(6) KEESOM, W. H. *Commun. phys. Lab. Univ. Leiden*, No. 71*d* (Supplement) (1932).
(7) WHITE, W. P. *J. Amer. chem. Soc.* **56**, 20 (1934); *Rev. sci. Instrum.* **5**, 269 (1934).
(8) THOMAS, J. L. *Bur. Stand. J. Res.*, Wash., **12**, 323 (1934).
(9) KEESOM, W. H., VAN DER HORST, H. and TACONIS, K. W. *Physica*, Eindhoven, **1**, 324 (1933–4).
(10) HEUSE, W. and OTTO, J. *Ann. Phys.*, Lpz., **2**, 1012 (1929).
(11) JACOBUS, D. D. Abstr. Sci. tech. Publ. 7, 54 (1931).
(12) MOSER, H. *Ann. Phys.*, Lpz., **14**, 790 (1932).
(13) BRUN, E. *C.R. Acad. Sci.*, Paris, **198**, 899 (1934).
(14) —— *C.R. Acad. Sci.*, Paris, **198**, 1213 (1934).
(15) ZETZMANN, J. *Arch. tech. Messen.* **4**, 116 (1934).
(16) HAASE, G. *Ann. Phys.*, Lpz., **20**, 75 (1934).
(17) DAVID, W. T. and JORDAN, J. J. *Phil. Mag.* **17**, 172 (1934); **18**, 228 (1934). See also DAVID, W. T. *Engineering*, **138**, 475 (1934).
(18) ELLIS, O. C. DE C. and MORGAN, E. *Trans. Faraday Soc.* **30**, 287 (1934).
(19) GRIFFITHS, E. and AWBERY, J. H. *Proc. roy. Soc.* **123**, 401 (1929).
(20) ROESER, W. F. and WENSEL, H. T. *Bur. Stand. J. Res.*, Wash., **12**, 519 (1934).
(21) EVANS, R. D. *Rev. sci. Instrum.* **4**, 391 (1933).
(22) LANGE, B. and VOOS, E. *Z. tech. Phys.* **15**, 323 (1934).
(23) CAMERON, A. E. *Rev. sci. Instrum.* **4**, 610 (1933).
(24) SCHATTENSTEIN, A. I. *Z. Elektrochem.* **40**, 653 (1934).
(25) KARSTEN, A. *Z. InstrumKde*, **54**, 274 (1934) (Thermostat).
(26) WOODWORTH, C. E. *Science*, **78**, 535 (1933) (Thermostat).
(27) VAN SANTEN, J. J. M. *Commun. phys. Lab. Univ. Leiden*, No. 227*a*, 1 (1934).

§ 2. CALORIMETRY

(28) SCHUKAREW, A. N., KRIVOBABKO, I. P. and SCHUKAREWA, L. A. *Phys. Z. Sowjet.* **5**, 722 (1934).
(29) SWIETOSLAWSKI, W. *Bull. int. Acad. Cracovie*, **1**, 64 (1934).
(30) SANIELEVICI, A. *J. Chim. phys.* **30**, 513 (1933).
(31) ZLOTOWSKI, I. *C.R. Acad. Sci.*, Paris, **199**, 284 (1934).
(32) TIAN, A. *J. Chim. phys.* **30**, 665 (1933).
(33) WHIPP, B. *Phil. Mag.* **18**, 745 (1934).

(34) ROBERTS, J. K. and WHIPP, B. *Proc. Camb. phil. Soc.* **30**, 376 (1934).
(35) JESSEL, R. *Proc. phys. Soc.* **46**, 747 (1934).
(36) KRAUS, C. A. and RIDDERHOF, J. A. *J. Amer. chem. Soc.* **56**, 79 (1934).
See also KRAUS, C. A. and SCHMIDT, F. C. *J. Amer. chem. Soc.* **56**, 2297 (1934).
(37) STEPHENS, R. W. *Phil. Mag.* **17**, 297 (1934).
(38) ASKEW, F. A., JACKSON, N. S., GATTY, O. and WOLFENDEN, J. H. *J. chem. Soc.* p. 1362 (1934).
(39) KEFFLER, L. J. P. *J. phys. Chem.* **38**, 717 (1934).
(40) JESSUP, R. S. and GREEN, C. B. *Bur. Stand. J. Res.*, Wash., **13**, 469 (1934).
(41) TAYLOR, G. I. and QUINNEY, H. *Proc. roy. Soc.* **143**, 307 (1934).
(42) FRANCE, R. W. *Trans. Faraday Soc.* **30**, 450 (1934).
(43) WOOD, W. A. *Phil. Mag.* **18**, 495 (1934).
(44) BECKER, G. and ROTH, W. A. *Z. Elektrochem.* **40**, 836 (1934). (Heats of formation.)

§ 3. CONDUCTION

(45) KANNULUIK, W. G. and MARTIN, L. H. *Proc. roy. Soc.* **144**, 496 (1934).
See also LABY, T. H. *Proc. roy. Soc.* **144**, 494 (1934).
(46) DICKINS, B. G. *Proc. roy. Soc.* **143**, 517 (1934).
(47) GREGORY, H. and ARCHER, C. T. *Phil. Mag.* **15**, 301 (1933).
(48) MILVERTON, S. W. *Phil. Mag.* **17**, 397 (1934).
(49) HERCUS, E. O. and SUTHERLAND, D. M. *Proc. roy. Soc.* **145**, 599 (1934).
(50) POWELL, R. W. *Proc. phys. Soc.* **46**, 659 (1934).
(51) VAN DUSEN and SHELTON, S. M. *Bur. Stand. J. Res.*, Wash., **12**, 441 (1934); **12**, 429 (1934).
(52) GRÜNEISEN, E. and REDDEMAN, H. *Ann. Phys.*, Lpz., **20**, 843 (1934).
See also REDDEMAN, H. *Ann. Phys.*, Lpz., **20**, 502 (1934).
(53) CINNAMON, C. A. *Phys. Rev.* **46**, 215 (1934).
(54) CRARY, A. P. *Physics*, **4**, 332 (1933).
(55) KISTLER, S. S. and CALDWELL, A. G. *J. industr. Engng Chem.* **26**, 658 (1934).
(56) VAN REST, E. D. *Proc. phys. Soc.* **46**, 537 (1934).
(57) GREEN, G. *Phil. Mag.* **18**, 625 (1934). (Theory of conduction.)
(58) LOWAN, A. N. *Phil. Mag.* **17**, 849 and 914 (1934). (Theory of conduction.)
(59) MUSKAT, M. *Physics*, **5**, 71 (1934). (Theory of conduction.)
(60) ROBERTSON, J. *Phil. Mag.* **18**, 165 and 1009 (1934). (Theory of conduction.)
(61) FOX, E. N. *Philos. Trans.* **232**, 431 (1934). (Theory of conduction.)

§ 4. RADIATION

(62) CARTWRIGHT, C. H. *Z. Phys.* **92**, 153 (1934).
(63) HARDY, J. D. *Rev. sci. Instrum.* **5**, 120 (1934).
(64) GUILD, J. *J. sci. Instrum.* **8**, 14 (1931); see also pp. 111 and 136.
(65) BOSWORTH, R. C. L. *Trans. Faraday Soc.* **30**, 554 (1934).
(66) HARRIS, L. *Phys. Rev.* **45**, 635 (1934).
(67) KOCH, W. *Z. tech. Phys.* **15**, 80 (1934).
(68) RITZOW, G. *Ann. Phys.*, Lpz., **19**, 769 (1934).
(69) HOFFMANN, F. and WILLENBERG, H. *Phys. Z.* **35**, 1 (1934).
(70) BUCKLEY, H. *Phil. Mag.* **17**, 576 (1934).
(71) CUNNOLD, F. A. and MILFORD, M. *Phil. Mag.* **18**, 561 (1934).
(72) FINKELNBERG, W. *Z. Phys.* **88**, 763 (1934).
(73) HULL, G. F., BELL, MARY and GREEN, S. E. *Proc. phys. Soc.* **46**, 589 (1934).
(74) CSÁSZÁR, E. *Z. Phys.* **90**, 667 (1934). (Generalization of Planck's law.)
(75) PLOTNIKOW, M. *Acta phys. polon.* **2**, 335 (1933).

§ 5. Convection

(76) Saunders, A. O. *Engineering*, **138**, 436 (1934).
(77) Ramdas, L. A. and Malurkar, S. L. *Nature*, Lond., **129**, 201 (1932).
(78) Hencky, K. and Neubert, P. *Naturwissenschaften*, **18**, 392 (1930).
(79) Kennard, R. B. *Phys. Rev.* **38**, 1921 (1931); *Bur. Stand. J. Res.*, Wash., **8**, 787 (1932).
(80) Hansen, G. *Z. tech. Phys.* **12**, 436 (1931).
(81) Schmid, E. *Forsch. IngWes.* **3**, 181 (1932); *Z. Ver. dtsch. Ing.* **76**, 1025 (1932).
(82) Townend, H. C. H. *J. sci. Instrum.* **11**, 184 (1934).
(83) Griffiths, E. and Awbery, J. H. *Proc. Instn mech. Engrs*, p. 319 (1933).
(84) King, W. J. *Mech. Engng*, N.Y., **54**, 190, 275, 296 and 347 (1932).
(85) Jodbauer, K. *Forsch. IngWes.* **4**, 157 (1933).
(86) Kaye, W. A. and Furnas, C. C. *J. industr. Engng Chem.* **26**, 783 (1934).
(87) Piercy, N. A. V. and Schmidt, R. J. *Phil. Mag.* **17**, 423 (1934).
(88) Jakob, M. and Fritz, W. *Forsch. IngWes.* **2**, 435 (1931).
(89) Jakob, M. and Linke, W. *Forsch. IngWes.* **4**, 75 (1933).
(90) Jakob, M. *Forsch. IngWes.* **3**, 102 (1932).
(91) Monrad, C. and Badger, W. L. *J. industr. Engng Chem.* **22**, 1103 (1930). (Critique and amplification of Nusselt's theory.)
(92) Hebbard, G. M. and Badger, W. L. *J. industr. Engng Chem.* **26**, 420 (1934).
(93) Kirkbride, C. G. *J. industr. Engng Chem.* **26**, 425 (1934).
(94) Houghton, H. G. *Physics*, **4**, 419 (1933).
(95) Fuchs, N. *Phys. Z. Sowjet.* **6**, 224 (1934).
(96) Jakob, M. *Phys. in regelmäss. Ber.* **1**, 153 (1933). (Survey of convection.)
(97) Chandler, R. *J. Instn Petrol. Tech.* **20**, 51 (1934). (Heat transfer.)
(98) Volmer, M. and Flood, H. *Z. phys. Chem.* **170**, 273 (1934). (Condensation.)
(99) Hilpert, R. *Forsch. IngWes.* **4**, 215 (1933). (An excellent paper on convection from wires and cylinders.)
(100) Jakob, M., Erk, S. and Eck, H. *Forsch. IngWes.* **3**, 161 (1932). (Condensation.)

§ 6. Vapour pressure

(101) Osborne, N. S. and Meyers, C. H. *Bur. Stand. J. Res.*, Wash., **13**, 1 (1934).
(102) Eck, H. *Forsch. IngWes.* **4**, 21 (1933).
(103) Rudberg, E. *Phys. Rev.* **46**, 763 (1934).
(104) Trautz, M. and Ader, H. *Phys. Z.* **35**, 446 (1934).
(105) Eucken, A. *Phys. Z.* **35**, 708 (1934).
(106) Verschaffelt, J. E. *Phys. Z.* **35**, 1013 (1934).
(107) See, for example, *Reports on Progress in Physics* (Physical Society), p. 177 (1934).
(108) Collins, E. M. *J. phys. Chem.* **37**, 1191 (1933). (Vapour pressure of solutions of sulphuric acid.)
(109) Lannung, A. *Z. phys. Chem.* **170**, 134 (1934).

§ 7. Liquids, vapours and gases

(110) Keesom, W. H. and Keesom, Miss A. P. *Commun. phys. Lab. Univ. Leiden*, No. 224*d* and 224*e* (1933) and 76*b* (1934).
(111) Keesom, W. H. *Proc. Kon. Akad. Wet. Amst.* **36**, 147 (1933).
(112) Keesom, W. H. and Keesom, Miss A. P. *Commun. phys. Lab. Univ. Leiden*, No. 221*d* (1932) or *Proc. Kon. Akad. Wet. Amst.* **35**, 736 (1932).
See also Keesom, W. H. *Commun. Phys. Lab. Univ. Leiden*, No. 217, Supplement 71*c* (1933).
(113) Halpern, O. *J. chem. Phys.* **1**, 880 (1933)
(114) Guggenheim, E. A. *Proc. Kon. Akad. Wet. Amst.* **37**, 294 (1934).
(115) Stewart, G. W. *Chem. Rev.* **6**, 483 (1929); *Rev. mod. Phys.* **2**, 116.
(116) Ward, H. K. *J. chem. Phys.* **2**, 153 (1934).

(117) KATZOFF, S. *J. chem. Phys.* **2**, 153 (1934).
(118) PARTHASARATHY, S. *Phil. Mag.* **18**, 90 (1934).
(119) GRINGRICH, N. S. and WARREN, B. E. *Phys. Rev.* **46**, 248 (1934).
(120) SPANGLER, R. D. *Phys. Rev.* **46**, 698 (1934).
(121) BENZ, C. A. and STEWART, G. W. *Phys. Rev.* **46**, 703 (1934).
(122) WINKLER, C. A. and MAASS, O. *Canad. J. Res.* **9**, 610 (1933).
(123) KOCH, W. *Forsch. IngWes.* **3**, 189 (1932).
(124) BIRGE, R. T. and JENKINS, F. A. *J. chem. Phys.* **2**, 167 (1934).
(125) HEUSE, W. and OTTO, J. *Phys. Z.* **35**, 57 and 628 (1934).
(126) COPPOCK, J. B. M. and WHYTLAW-GRAY, R. *Proc. roy. Soc.* **143**, 487 (1934).
(127) STEWART, W. W. and MAASS, O. *Canad. J. Res.* **11**, 530 (1934).
(128) KEESOM, W. H. and VAN SENTEN, J. J. M. *Commun. phys. Lab. Univ. Leiden*, No. 227*b* (1933). (Isothermals of helium.)
(129) MACFARLANE, W. and WRIGHT, R. *J. chem. Soc.* p. 207 (1934). ("Solubility" of vapours in gases.)
(130) FERGUSON, A. and MILLER, J. T. *Proc. phys. Soc.* **46**, 140 (1934). (Formula for molecular volume as function of temperature.)
(131) JOHNSTON, H. L. and WEIMER, H. R. *J. Amer. chem. Soc.* **56**, 625 (1934). (Characteristic surface for oxides of nitrogen.)
(132) VAN ITTERBECK, A. *Commun. phys. Lab. Univ. Leiden*, No. 217 (1933). (Properties of helium.)
(133) ROEBUCK, J. R. and OSTERBERG, H. *Phys. Rev.* **45**, 332 (1934). (Properties of helium.)

§ 8. HYGROMETRY

(134) WILLIAMS, S. R. *Rev. sci. Instrum.* **4**, 549 (1933).
(135) MOSS, E. B. *Proc. phys. Soc.* **46**, 450 (1934).
(136) AWBERY, J. H. and GRIFFITHS, E. *Proc. phys. Soc.* **44**, 132 (1932).
(137) WHIPPLE, F. J. W. *Proc. phys. Soc.* **45**, 307 (1933).
(138) WENK, F. *Z. InstrumKde*, **54**, 15 (1934).
(139) GRUNDMANN, W. *Z. InstrumKde*, **54**, 441 (1934).
(140) SCHWEITZER, E. *Naturwissenschaften*, **21**, 784 (1933). (Cellophane in hygrometry.)
(141) BROOKS, D. B. *Bur. Stand. U.S. Misc. Publ.* No. 143 (1933). (Psychrometric chart.)
(142) HUMPHREYS, W. J. *Mon. Weath. Rev.*, Wash., **61**, 300 (1933).

§ 9. THERMODYNAMICS

(143) LEWIS, B. and VON ELBE, G. *J. chem. Phys.* **2**, 659 (1934).
(144) MAXWELL, G. B. and WHEELER, R. V. *J. chem. Soc.* p. 882 (1933). For the opposite view, compare reference (17).
(145) EASTMAN, E. D. and MILNER, R. T. *J. chem. Phys.* **1**, 444 (1933).
(146) RHODEBUSH, W. H. *J. chem. Phys.* **2**, 668 (1934). See also reference (159).
(147) KIRKWOOD, J. G. *J. chem. Phys.* **2**, 767 (1934).
See also HALPERN, O. *J. chem. Phys.* **2**, 85 (1934).
(148) BEEBE, R. A. and WILDNER, E. L. *J. Amer. chem. Soc.* **56**, 642 (1934).
(149) ZICKERMANN, C., DRAKE, L. C., BENTON, A. F. and HINSHELWOOD, C. H. *Z. Phys.* **88**, 43 (1934).
See also *J. Amer. chem. Soc.* **56**, 506 (1934); *J. chem. Soc.* p. 1357 (1933) and reference (34).
(150) DONNAN, F. G. *Z. phys. Chem.* A, **5**, 369 (1934).
(151) GATTY, O. *Phil. Mag.* **18**, 273 (1934).
(152) ROEBUCK, J. R. and OSTERBERG, H. *Phys. Rev.* **46**, 785 (1934). (Joule-Thomson effect for argon.)
(153) CRAXFORD, S. R., GATTY, O. and PHILPOTT, J. ST L. *Phil. Mag.* **17**, 54 (1934). (Thermodynamics of surface tension.)
(154) BONE, W. A. and LAMONT, F. G. *Proc. roy. Soc.* **144**, 250 (1934). (Thermodynamics of combustion.)

§ 10. Specific heats of gases

(155) Lewis, B. and von Elbe, G. *J. chem. Phys.* **2**, 291 (1934); **2**, 294 (1934); but see **2**, 890, (1934).
(156) Johnston, H. L. and Long, E. A. *J. chem. Phys.* **2**, 389 (1934). (A correction on p. 710.)
(157) Farkas, A. and Farkas, L. *Trans. Faraday Soc.* **30**, 1071 (1934).
(158) Zeise, H. *Z. Elektrochem.* **39**, 758 (1933); **39**, 895 (1933); **40**, 662 (1934) (corrections).
(159) Clusius, C. *Z. Electrochem.* **40**, 98 (1934).
(160) Gordon, A. R. *J. chem. Phys.* **2**, 65 (1934).
(161) Sherratt, G. G. and Griffiths, E. *Proc. roy. Soc.* **147**, 292 (1934).
(162) Eucken, A. and Becker, R. *Z. phys. Chem.* B, **27**, 235 (1934).
See also Richardson, E. G. *Proc. roy. Soc.* **146**, 56 (1934).
(163) David, W. T. and Leah, A. S. *Phil. Mag.* **18**, 307 (1934).
(164) Renner, R. *Phys. Z.* **35**, 811 (1934). (Specific heat of methane.)
(165) Van Itterbeck, A. *Commun. phys. Lab. Univ. Leiden*, No. 217 and 218, Supplement 70*b*. (Hydrogen under pressure.)
(166) Cassel, L. S. *J. Amer. chem. Soc.* **56**, 1838 (1934). (Nitrous oxide.)

§ 11. Dilatation

(167) Owen, E. A. and Yates, E. L. *Phil. Mag.* **17**, 113 (1934).
See also Owen, E. A. and Pickup, L. *Proc. roy. Soc.* **145**, 258 (1934).
(168) Jesse, W. P. *Physics*, **5**, 147 (1934).
(169) Jay, A. H. *Proc. roy. Soc.* **143**, 465 (1934).
(170) Buchta, J. W. and Goetz, A. *Phys. Rev.* **46**, 1092 (1934).

§ 12. Miscellaneous

(171) Henning, F. *Phys. in regelmäss. Ber.* **2**, 1 (1934). (Survey of the whole subject.)
(172) Griffiths, R. *Thermostats* (1934). London: Griffin. (Book on the subject.)
(173) Trautz, M. and Zündel, F. *Z. techn. Phys.* **12**, 273 (1931). (Survey of conduction data.)
(174) Ebert, H. *Phys. in regelmäss. Ber.* **2**, 123 (1934). (Survey of hygrometry.)

ELECTRICAL AND MAGNETIC MEASUREMENTS

By L. HARTSHORN

In the following report no attempt has been made to deal with all the aspects of the subject in which progress has been made, or even to single out those which are considered to be the most important. The plan adopted has been to select a few topics which seem to have attracted a good deal of attention during the year, and to record their present positions, linking these up with the work of the few preceding years where this seemed advisable. It is hoped that important items which have found no place in this report or its predecessor, will in due course find their way into subsequent reports of the series.

§ 1. THE MEASUREMENT OF VOLTAGE

Electrometers of high sensitivity. There has been a pronounced tendency in recent years to use thermionic valves for the measurement of very small potential-differences (cf. last year's Report), and, since the invention of the well-known Lindemann electrometer, the amount of work carried out with the object of providing better electrometers of the ordinary type has been relatively small. Mention should however be made of Perucca's instrument which has much in common with Lindemann's, and also the recent work of Carmichael, which has resulted in a greatly improved form of the Wilson tilted gold-leaf electrometer. Perucca's[1] instrument has a very light "needle" consisting of a quartz-fibre bent into the form of a U, attached at points near its open ends to a quartz torsion-fibre, which is in tension. Torsion thread and needle are metal-coated. The needle moves between two fixed plates, to which the auxiliary p.d. is applied. The tip of the quartz-fibre needle, which is made very thin, is observed by means of a microscope. A sensitivity of 1 mV. per division, with the movement aperiodic, and a time of indication of 5 sec. may be obtained.

Carmichael[2,3] has described two forms of the tilted electrometer. In the earlier one[2], the gold leaf was replaced by a silvered quartz-fibre suspended by a hinge of gold leaf, and the plate was made very narrow so that the movements of the fibre were confined to one plane. The instrument was almost completely metal-clad and the case made gas-tight, so that there was complete protection from thermal and atmospheric disturbances. The necessary adjustments were very simply made, and it was found that the sensitivity which could be used was only limited by the extreme slowness of the deflections at very high sensitivity due to excessive air-damping. Considerable improvement in this respect was obtained by filling the instrument with hydrogen, and a sensitivity of 1 mV. per division was satisfactorily used.

In a subsequent investigation[3] the air-damping was greatly reduced and controlled by the evacuation of the instrument. The time of response was greatly diminished and the movement could be made critically damped. It was then found that the chief limitation to the working sensitivity was set by imperfections in the elastic properties of the fibre, which lead to irregularities in the calibration curves, a slow "creep" of the deflections (elastic hysteresis or after-effect), and instability of the zero arising from undue sensitivity to slight changes of level or shape of the leaf. The best results were obtained with the thinnest available quartz-fibre (about $1{\cdot}5\,\mu$ diameter) very lightly sputtered with gold. No gold-leaf hinge is necessary as the fibre itself is more flexible than the gold leaf. The metal coating of the fibre was found to be the cause of most of the imperfections, which disappeared when the coating was made extremely thin. The sensitivity obtained in this way by the use of very fine fibres, in a vacuum adjusted for critical damping, was limited only by the Brownian motion of the fibre, so that the maximum possible sensitivity for this type of instrument was reached. The minimum potential-change that could be detected was of the order 1×10^{-4} volt when the period (undamped) was 5 sec., and it was approximately inversely proportional to the period. The radiometer force on the fibre in the evacuated electrometer was detectable, but did not affect the performance of the instrument so long as the gas-pressure and intensity of illumination of the fibre were not varied. The limit of sensitivity set by Brownian movements had previously been reached by Eggers[4] with a Hoffmann[14] electrometer, but it is pointed out that the tilted electrometer possesses the great advantage that the moment of inertia of its hanging fibre may be made smaller than that of any suspended "needle" such as is used in the various torsion electrometers, so that a high sensitivity is obtained with a shorter period. The capacitance of the instrument is very small and the sensitivity to charge correspondingly high. It was shown that the ionization due to single α particles could be measured.

Electrometers for measurements at very high frequencies. Alternating currents of frequencies of the order of 10^6 to 10^8 c./sec. are now very widely used for physical measurements, and there have been many investigations of methods of measurement of voltage at these frequencies. The measuring instrument must have a high impedance, and at high frequencies this can only be secured by making its capacitance very small. Thus the double-thread string electrometer of Wulf[5] suggests itself as likely to be the most satisfactory instrument for very high frequencies, owing to its simplicity, and the smallness of its capacitance. In the ordinary form of the instrument one electrode consists of two very fine Wollaston wires, parallel and very close together, their upper ends connected to some form of terminal, and their lower ends attached to a bent quartz-fibre, which keeps them in tension. The other electrode is a metal cylinder surrounding the wires. When the potential of the wires differs from that of the outer cylinder, they become charged and repel one another. The movements of their central portions are observed by means of a microscope, and serve to measure the potential-difference. When the instrument is used for a.-c. work it is traversed by a displacement current proportional to the voltage, the frequency, and the capacitance, and at frequencies of the order of 10^8 this current

may well amount to several milliamperes. An equivalent conduction current must flow through the wires, and since they must be very thin in order to obtain high sensitivity, errors of two kinds may arise. (1) Errors due to displacements of the wires caused by the heating effect of the current. (2) The potential along the wires will vary owing to their resistance and inductance, and the deflection will no longer measure the terminal potential-difference, but that between the middle portions of the wires, which will be larger. G. Ulbricht[7] has obtained equations for the second of these errors by applying the equations for a transmission line to the wires. He estimates that the error with the Wulf instrument is less than 1 per cent at a frequency of 10^8 c./sec., but increases very rapidly beyond this point. H. F. Nissen[6] had modified the instrument for high-frequency work. The wires are replaced by short thin strips of silver or aluminium, which are extended by insulating threads attached to a bent quartz-fibre as in the original instrument. The strip form of electrode gives lower resistance and inductance, which will tend to reduce the above-mentioned errors. The short length of the strips keeps the capacitance low ($0{\cdot}6\,\mu\mu$F. rising to $0{\cdot}65\,\mu\mu$F. with the deflection), while the insulating extensions allow the sensitivity to be maintained. It is the deflection of the insulating threads which is observed, since these are very thin and therefore suitable for microscopic observation. Since the strips or wires carry current, there may also be an error arising from the electromagnetic force between them, but Nissen estimates that all the errors are entirely negligible at frequencies up to 5×10^7 c./sec. and that the instrument is satisfactory up to $2{\cdot}10^8$ c./sec. Perhaps the most serious limitation is the fact that for any measurements, a certain length of lead must be connected to the instrument, and owing to the inductance of the lead and the capacitance of the instrument, the instrument terminal-voltage will usually exceed the voltage to be measured. This error is considered negligible so long as the total length of the system, leads + instrument, is less than one-twentieth of the wave-length, and with the shortest practical length of lead the error is likely to be of the order of 1 per cent at $2{\cdot}10^8$ c./sec. The range of the instrument is from 10 to 80 volts, the scale being very nearly linear over the range 20–80. The range can in many cases be extended to very high voltages by means of potential-dividers. These instruments possess the great advantages that they are simple and compact, require no auxiliary sources of current or voltage, and may be calibrated by d.-c. methods.

Voltmeters for high-frequency work. For a great deal of work at frequencies up to 10^7 c./sec. it is possible to use instruments with a capacitance considerably larger than that of the electrometers so far considered. The available deflecting force is then larger and the instrument may be given a more convenient form. A. Palm[8] has described a portable electrostatic voltmeter, with a scale which is linear from 10 to 160 volts, the capacitance varying from 11 to 13 $\mu\mu$F. with the deflection. It is a pointer instrument and the wire forming the pointer also forms the moving electrode. The suspension is of thin metal strip. The air-damping is complete and the instrument takes up its reading in 15–20 sec. Portable instruments of somewhat similar construction, but with ranges from 125 to 1000 volts (capacitances 54 and 17 $\mu\mu$F. respectively) are described by A. Täuber-Gretler[9]. A reflecting instru-

ment with somewhat similar properties but suitable for the range 500–2500 volts has been described by B. A. Sharpe[10]. This instrument is oil-damped, and in order to avoid the passage of current through the suspension, the moving element carries a steel needle dipping into a mercury cup, which serves as one terminal of the instrument. The difficulties associated with the current in the suspension-fibre are avoided in a novel design due to A. Bestelmeyer[11]. Both terminals of the instrument are connected to fixed electrodes which form a cylindrical metal box, closed except for narrow cuts, necessary for the insulation of the two parts. A light aluminium needle is suspended inside this box by means of a quartz-fibre, so that it is completely insulated from both terminals. When voltage is applied to the fixed electrodes, the needle tends to set itself so as to make the overall capacitance of the instrument as large as possible. The potential of the needle will vary with the deflection, and will depend only on the displacement currents flowing to and from it. Evidently perfect shielding of the needle will be of the utmost importance. The full scale reading is stated to be 20 volts and the scale is linear.

In all these instruments the deflecting force may be represented by $V^2\, dC/d\alpha$, where V is the applied potential, C the capacitance of the instrument, and α the deflection. The restoring force due to the suspension is proportional to α, so that a linear scale is obtained by making $dC/d\alpha \propto 1/\alpha$. In Palm's instrument the moving element is a straight wire and the fixed elements sickle-shaped. Sharpe's instrument is based on a previous design due to Wilkinson[12] in which the moving element consists of two spheres joined by a straight wire, and the fixed elements consist of circular rings which are approached by the moving spheres. Bestelmeyer's electrodes are less regular in shape.

An obvious way of extending the range of the low-reading voltmeters is the use of capacitance potential-dividers. A simple divider consisting of three insulated parallel plates, suitable for voltages up to 60,000 and frequencies up to 10^6 c./sec., has been described by Richardson[13].

Three-terminal electrometers for alternating potentials. The instruments so far considered for the measurement of alternating potentials are all of the two-element type, and these are very much less sensitive than those of the three-element type, such as the quadrant electrometer, which require the use of an auxiliary voltage. Investigations concerning the use of this type of electrometer for alternating potentials at frequencies up to 10^6 c./sec. have been made, and a sensitivity of 0·2 mV. per mm. deflection at 1-metre scale distance has been obtained, but it would be a mistake to imagine that any desired voltage can be measured with anything like this accuracy. It is necessary to have an auxiliary voltage of the order of 40 volts, of the same frequency and phase as that to be measured, and no general means of obtaining this has been devised. The experiments made do, however, demonstrate the possibility of using these sensitive electrometers for certain comparative measurements, such as bridge measurements. The work published refers chiefly to Hoffmann's[14] "duant" electrometer, which may be roughly described as half a quadrant electrometer made on a small scale. A cylindrical box forming the fixed electrodes is divided into two "duants" instead of the more usual four quadrants, and the needle

is of sector form, with the suspension attached near the pointed end, the mirror serving as a counter-weight. The sensitivity may be increased either by bringing the needle nearer to the duants, or by increasing the potential-difference between them. A condition of maximum sensitivity may be obtained by either adjustment. In its most sensitive form for d.-c. work the instrument is made so that the needle and duants are in a vacuum, and extremely well shielded from thermal disturbances, but for a.-c. work these refinements are dispensed with and the electrodes are on a somewhat larger scale, e.g. the silver needle is 1–2 cm. long, suspended by Wollaston wire of 6μ diameter. For low-frequency work[15] the auxiliary potential applied to the duants is obtained from a resistance potential-divider, the case of the instrument being connected to its middle point and to earth. The potential to be measured is applied between the case and the needle. Provided there is no phase difference between the potentials of the elements of the instrument, the ordinary equations obtained by electrostatic theory must be used. The necessary condition of phase equality is easily satisfied at low frequencies by obtaining all the potentials from a common source, by means of resistance potential-dividers; but at frequencies up to 10^6 c./sec. this is no longer satisfactory, as the phase changes in a potential-divider of this form are approximately proportional to the frequency. In the measurements of C. Michel[16] the auxiliary p.d. is applied to the duants by connecting them to an inductive coil coupled to the oscillator. The needle and case of the instrument are permanently connected together and to earth; and the p.d. to be measured is applied between their common terminal, and the mid-point of the coupling coil connected to the duants. Thus this p.d. disturbs the equality of the p.d. between duants and needle, and thereby causes a deflection. The phase of the duant potentials is adjusted by means of a variable inductance in the leads connecting them to the coupling coil. It is obvious that the necessary conditions will not be easily satisfied, but the experiments show that it is possible to obtain the high sensitivity previously quoted. Similar considerations must also apply to any sensitive three-electrode instrument, such as the Compton, Lindemann and Perucca electrometers, and the ordinary single-string instrument, but for high-frequency work it is important that the impedance of each electrode shall be low, so that an instrument with a metal-coated quartz-fibre as one electrode would probably not be suitable. The theory of electrometers used in this way at high frequencies has been discussed by A. Hermann[17].

Peak voltmeters. It is frequently desirable to be able to measure the maximum value of an alternating voltage, and there have been several recent discussions of instruments for this purpose, which are usually called peak voltmeters. A simple form of this instrument consists of a rectifier in series with a condenser, with an electrostatic voltmeter connected so as to measure the voltage across the condenser. When the voltage applied to the series-combination of rectifier and condenser is in such a direction that the rectifier passes current, the condenser is charged, and the p.d. across it rises until it is equal to the maximum value of the applied voltage, after which the flow of current ceases. When the applied voltage falls, current in the reverse direction is stopped by the rectifier. Any leakage in the electrostatic voltmeter will result in a loss of charge by the condenser, and tend to give too low a value. The

condenser should be chosen so that the time-constant of the circuit consisting of its capacitance with the total conductance in parallel with it, is a maximum, and a condenser of large capacitance is therefore desirable. The time-constant of the capacitance in series with the forward resistance of the rectifier should however be small so that the condenser will reach its maximum charge fairly quickly. The "reverse" conductance of the rectifier must be very low, for this also is equivalent to conductance in the condenser. Banner[18] points out that the copper-oxide rectifier is not suitable for such work, since its reverse resistance is too small. The thermionic valve is however quite satisfactory, and it is convenient to use an ordinary triode as a diode, by connecting the grid to the anode. The range of such a voltmeter depends on that of the electrostatic voltmeter used for the measurement of the constant p.d. to which the condenser becomes charged. It is pointed out by Sims[19] that since the condenser p.d. is constant, the range of the instrument may be made either larger or smaller by connecting a battery in series with the electrostatic voltmeter, so that the p.d. of the battery is either subtracted from, or added to that of the condenser. An alternative means[20] of obtaining an instrument of low range (1–10 volts), is to use a triode for the measurement of the p.d. of the condenser. If the triode is connected so that its grid becomes negatively charged, the leakage current is not very serious provided that a large condenser (2 μF.) is used. The readings taken are those of a milliammeter in the anode-circuit of the triode. When such an instrument was calibrated with a sine-wave voltage an accuracy of 2 per cent was obtained in subsequent measurements on various wave-forms.

When very high peak voltages are to be measured, say 10–100 kV., the measurement of the p.d. to which the condenser becomes charged is no longer a simple matter, and it is more convenient to use two rectifiers arranged so that the condenser is charged to the maximum voltage in one direction through one rectifier, and then discharged and charged to the maximum voltage in the other direction through the other rectifier; and so on with every alternation of the applied voltage. Evidently the mean current i_m through either rectifier will be proportional to the maximum voltage V_m, the frequency f and the capacitance of the condenser C; and the peak voltage may be simply determined from the equation $i_m = 2fCV_m$. The two rectifiers must be in parallel, but must act in opposite directions. In early forms of the instrument mechanical rectifiers were used, but the necessary condition that the contacts make and break simultaneously so that one and only one pair is closed at any moment is difficult to satisfy, and is only possible when the wave is periodic and vanishes in equal intervals of time, i.e. when there are no even harmonics present. This restriction is overcome when two diodes are used as rectifiers, but errors then arise from the fact that diodes have an impedance which may not be negligible, and that they give an emission current of about 2 per cent of their normal value, even when no voltage is applied to the anode. Thus current circulates round the two rectifiers in parallel, and affects the measured mean current, so that the above equation is not strictly accurate. This emission current may however be stopped by a bias-battery, and A. T. Starr[21] has shown that this arrangement not only avoids the error due to the emission current at zero voltage, but also diminishes that due to the im-

pedance of the rectifier. Moreover, the adjustment of the bias-voltage is not critical, and provided the bias exceeds the critical value for zero current by an amount which is small compared with the voltage to be measured, its variation with time, and also those of the valve characteristics, are of no importance; and the instrument may be used as a standard without the necessity of continual re-calibration. The mathematical theory of the various errors is discussed.

The rotary voltmeter. A form of voltmeter which is novel, but which is in some respects analogous to the peak voltmeter employing mechanical rectifiers, has been described by Kirkpatrick and Miyake[22]. It was first devised for the measurement of constant voltages[23] from 1 to 100 kV. A brass cylinder divided longitudinally into two halves rotates at constant speed about its axis in the electrical field between two fixed electrodes, which may be regarded as the terminals of the instrument. They may be spheres, discs, or half-cylinders coaxial with the rotor, depending on the range of the instrument. The two halves of the rotor are connected together by leads which include a d.-c. galvanometer and a simple rectifying commutator of two sections, which is mounted on the same spindle as the rotor. Commutation occurs at the instant at which the capacitance between either section of the rotor and the nearest fixed electrode is a maximum. As the rotor rotates, each half becomes alternatively positively and negatively charged, the signs of their two charges at any instant being opposite. There is therefore a flow of electricity to and fro along the conductors joining the rotor-sections, and as commutation occurs at the instant at which this flow changes its direction, the current in the galvanometer is rectified, and the deflection gives its mean value. The rotor and fixed electrodes are symmetrically arranged, and it is easy to see that if V is the voltage between the fixed electrodes, and C is the maximum change of capacitance between either section of the rotor and one of the fixed electrodes during a revolution, the total transfer of charge in each half-revolution is VC, and the mean current in the galvanometer is therefore given by $i_m = 2nVC$, where n in the number of revolutions per sec. Thus for a fixed speed of rotation (which is easily obtained by the use of a synchronous motor) the current i_m determines the voltage V, and the galvanometer may be calibrated to read volts directly. The method has some very useful features, viz. a linear scale; a very wide range which is easily adjustable; and the possibility of taking readings at a distance. The potential of the rotor system is immaterial, and one point on it may therefore be earth-connected.

The instrument has also been adapted for use with alternating voltages[24], and it has been shown that it may be used for the delineation of wave-forms by a point-to-point method. For the measurement of constant voltages the rotor and commutator may be driven at any desired speed, but for a.-c. work the speed must be such that a half-revolution of rotor and commutator occurs in the same time as one or more complete cycles of the voltage, so that the instantaneous voltage whenever commutation occurs is always the same. The net transfer of charge in the interval between two successive instants of commutation depends only on the initial and final values of V and C; and since the galvanometer reading depends only on the net transfer of charge and the speed, the reading is the same as if the voltage were constant, at the

above-mentioned instantaneous value. If the phase of commutator and rotor may be adjusted with respect to that of the voltage to be measured, the instantaneous voltage at various points on the wave may be measured, and so the wave-form obtained. The rotor is driven by a synchronous motor and the necessary phase adjustment made either by rotating the motor itself (a special form of mounting being devised for this purpose) or by supplying it with current from a phase-shifting transformer, the phase of which is changed as required.

§ 2. VOLTAGE STABILIZATION

A device has been developed in the course of the last few years which promises to be extremely useful in experiments in which it is necessary to maintain a constant voltage for long periods. It depends on the fact that when a glow-discharge is maintained between two electrodes in a rarefied gas, the voltage between the electrodes varies but little with changes in the value of the discharge-current. If therefore a discharge-tube is connected in series with a fixed resistance to the ordinary supply-terminals, the fluctuation of the supply-voltage will produce variations of the current, but the voltage between the electrodes of the tube will be nearly constant, and any apparatus connected to these electrodes will have a constant voltage applied to it. With the commercial forms of the apparatus now available, it is stated that the output-voltage does not vary more than 0·2 per cent, when the supply-voltage varies by 10 per cent. The fluctuations of the supply-voltage are borne almost entirely by the series-resistor, and this must have such a value that the corresponding changes of current are within the current-carrying capacity of the discharge-tube and associated apparatus. It will evidently be an advantage if the voltage between the terminals of this resistor changes rapidly for small changes of current. A barretter (an iron wire in a bulb containing hydrogen) is therefore used.

In its commercial form the device (Stabilovolt tube, stabilizator, or glow-discharge potential-divider) consists of a glass bulb, similar to that of a thermionic valve, filled with a gas (chiefly neon) at a pressure of a few cm. of mercury, and provided with a number of iron electrodes in the form of concentric cylinders closed at one end. The discharge passes across the gaps between the electrodes in series, and it is possible to obtain constant voltages of various amounts, by making connection to an appropriate pair of electrodes. The voltage required to start a discharge across any gap is greater than that required to maintain it, and it is therefore necessary to connect each of the inner electrodes to one of the main supply-terminals by means of a high resistance, so that the full supply-voltage is applied to each gap in turn in starting the discharge. The principle of stabilization appears to have been devised by F. Schröter a few years ago, while L. Körös introduced the multiple-gap potential-divider. Its basic properties[1] and the methods of calculation of the best working conditions[2] are described in papers by Körös and Seidelbach. An account in English[3] is also available. K. Lämmchen[4] has shown that if the discharge-tube forms one arm of a Wheatstone Bridge network, an even closer approximation to a constant voltage may be obtained from the "galvanometer terminals" of the network, while H. Meyer[5] has shown that a similar arrangement with two tubes in

opposite arms of the bridge is better still. It may be remarked that the last arrangement resembles the Callendar hot-wire current-regulator[6].

When the current required is only very small, the thermionic valve provides an alternative means of voltage stabilization. Several arrangements have been described for supplying constant voltages to electrometers, ionization chambers, etc., and it is stated that an output up to several thousand volts, in which the fluctuations are less than 1 part in 5000, may be obtained from an ordinary mains supply in this way. A valve with two or more grids (tetrode or pentode) is required. The potentials of the grids are obtained from a resistance potential-divider, which is connected across the fluctuating d.-c. supply. Thus the potentials of both grids fluctuate together and by proportional amounts, and by arranging the circuit so that they act upon the anode current in opposite directions, this current may be made independent of the fluctuations of the supply-voltage. The required constant voltage is then obtained from a potentiometer of high resistance which is included in the anode circuit of the valve. A detailed discussion of such circuits has been given by R. D. Evans[7] following earlier work of J. C. Street and T. H. Johnson[8]; and Richards[9].

§ 3. THE MEASUREMENT OF CURRENT

(1) *Absolute measurement.* At a recent General Conference on Weights and Measures it was decided that an absolute unit of current should be adopted for general use in the place of the present international unit, which is defined in terms of the electro-deposition of silver. This decision means that in future the electrodynamometer or current balance is to be regarded as the standard instrument for the measurement of current instead of the silver voltameter. Considerable interest therefore attaches to absolute measurements of current now in progress in several national standardizing laboratories. An account[1] of measurements recently made in America was published during 1934. The balance was that used by Rosa, Dorsey, and Millar in 1911, although it has been rebuilt and modified in several details. Essentially it is a Rayleigh balance, a moving coil with its plane horizontal being suspended from one pan, midway between two fixed coils, coaxial with it, and placed at such a distance from it that the force between fixed and moving coils is a maximum. Neglecting small corrections, the only quantity to be measured in addition to this force is the ratio of the radii of the coils (assuming the numbers of turns to be known). This ratio is determined electromagnetically so that no first-order precision measurements of linear dimensions are necessary, and it is possible to use coils of many layers and turns so as to increase the force to be measured. In order to measure the ratio of the radii, the coils are mounted so as to be concentric and coplanar, and with their common plane vertical and in the magnetic meridian. A small magnet is suspended at the centre, and currents through the two coils in opposition are adjusted until there is no effect on the needle. The ratio of the currents then gives the ratio of the radii.

The simple theory is only strictly applicable when the dimensions of the cross-sections of the coils are negligible in comparison with their radii, and in practice

this does not hold. Snow[2] has therefore worked out in considerable detail the force between two parallel coaxial coils whose windings have rectangular cross-sections. The ratio of cross-sectional dimensions to radius is assumed to be small, but only terms of the sixth and higher orders are neglected. The effects of the non-uniform distribution of current over the cross-section are considered, but it is shown that these are eliminated when the ratio of the radii of the coils is determined electromagnetically in the way outlined above.

As in all recent investigations of this kind the results were recorded in terms of resistance standards and standard cells, i.e. the current passing through the balance was also passed through a resistance which was varied until the p.d. across its terminals was equal to the e.m.f. of a standard cell. Indeed such a combination of standard cell and resistance is the only really practical standard of current. It may be calibrated in international units by comparison with the silver voltameter, and in absolute units with the current balance, and so the ratio between these units found. Unfortunately there is usually a gap of months or years between the two sets of measurements, although the evidence of these measurements and those of 1911 suggests that standard cells and resistors are reproducible or constant to the accuracy of the current measurements over a period of several years.

The results obtained with the American current balance using various combinations of coils showed a maximum deviation of eight parts in 10^6. It is interesting to note that the most recent silver voltameter measurements[3,4,5] showed a maximum deviation of about three times this amount, although the accuracy of the final result was in both cases estimated at 2×10^{-5}.

(2) *Galvanometers.* There can be little doubt that the ordinary moving-coil galvanometer is, within its range of sensitivity, by far the most convenient instrument for the measurement of small currents, and the conditions for high sensitivity and quick response in these instruments are of considerable importance. Prof. A. V. Hill[6] has considered the theory of the instrument from the practical standpoint, and has given the following convenient equation for its performance when critically damped:

$$P^2 = 3{\cdot}1\, MK^2/RT^2\,(1-\lambda/\pi),$$

where P is the μA. per mm. deflection at 1 m., R the total circuit resistance in ohms, MK^2 the moment of inertia of coil in g.cm^2, T the undamped period, and λ the logarithmic decrement on open circuit, i.e. a measure of the air-damping of the moving system.

The performance of galvanometers of various types is considered in relation to this equation, and the importance of a small coil (low MK^2) and the minimum of air-damping (small mirror, etc.) is pointed out.

(3) *Galvanometer amplification.* The output of papers on devices for increasing the sensitivity of the moving-coil galvanometer continues to be very large. An outline of the work on thermal relays or amplifiers, and on amplifying circuits employing thermionic valves was given in last year's *Report*. The more recent work on these items is mainly concerned with details such as the choice of valves and other circuit components. The papers by Distad and Williams[7] and Donzelot and

Divoux[8] may be quoted as typical for the benefit of those who require details of circuits for the measurement of currents as small as 10^{-14} ampere. E. Hasché[9] advocates a circuit employing a differential galvanometer.

The possibilities of purely optical amplification are considered by Dubar[10], and he has shown how the sensitivity of the ordinary lamp and scale arrangement may be increased by placing an auxiliary fixed mirror so as to obtain multiple reflection of the beam of light from the moving mirror. The angle of rotation of the emerging beam is $2n$ times that of the moving mirror, where there are n reflections from the latter. By the use of a fixed convex mirror the effective scale distance may be made large, while keeping the whole apparatus compact. The magnification obtainable in this way is of course limited by lack of definition due to diffraction and by loss of light.

The thermionic amplifier is chiefly used for the measurement of current in circuits of very high resistance. In circuits of low resistance, e.g. with thermocouples, the voltages available are very much smaller, and valve-amplification is correspondingly less effective. The thermal relay appears to have been hitherto the most successful device for circuits of low resistance, but recent developments suggest that the photo-electric relay may prove to be better.

(4) *The photo-electric relay.* It is now well known that a layer of cuprous oxide formed on copper possesses photo-electric as well as rectifying properties. The use of a photo-electric cell of this type for the magnification of galvanometer deflections was first suggested by A. V. Hill[11] in 1931. He showed that with two good galvanometers and such a cell, the limit of sensitivity set by Brownian movements could be reached, but more sensitive cells[12] are now commercially available, and a more efficient method of using the cell has been devised[13] so that the limiting sensitivity is now obtainable with quicker response, and it is this which gives the method its peculiar value. The cell is cut into two halves, which are connected in parallel, but in opposition to the secondary galvanometer. Thus when the two halves are equally illuminated, there is no resultant current through the galvanometer; but when the illumination of the halves is unequal there will be a current in one direction or the other, depending on which half receives the more light. The current passed by a cell is directly proportional to the total amount of light falling on its sensitive surface. The spot of light obtained by reflection from the mirror of the primary galvanometer in its zero position is arranged to fall symmetrically on the two halves of the cell so that the secondary galvanometer is also at its zero. Any deflection of the primary galvanometer then causes unequal illumination of the two halves of the cell, which produces a similar but much larger deflection of the secondary galvanometer. The response of the cell is practically instantaneous, and this gives it a great advantage over the thermal relay; for it is important that the overall response-time of the equipment shall be as small as possible, especially when varying currents are to be observed. A. V. Hill[14] has shown how the time of response of a combination of thermopile and galvanometer depends on the separate response-times of the thermopile and galvanometer. He has also considered the case of two galvanometers coupled by a photo-electric cell[15] and points out that in this case it is feasible to use

a secondary galvanometer with a period of only 0·2 sec., in which case the movements of the primary galvanometer are magnified with a lag of only 0·15 sec., whereas with the thermal relay the lag is likely to be 2 sec. or more.

R. V. Jones[16] has given practical details concerning the fitting of a photo-electric relay to a Moll galvanometer, so as to obtain a sensitivity of 10^{-11} amp. per mm., at which sensitivity the zero variations due to Brownian movements amount to 0·5 mm.

§ 4. THE ELECTRICAL CONDUCTIVITY OF THIN METALLIC FILMS

A considerable amount of work has been published during 1934 on the electrical conductivity of metals in the form of very thin films, such as are usually prepared by sputtering or evaporation. The data obtained are very complicated, and in some instances apparently contradictory, but some of the properties are more or less confirmed by all workers in this field, and it is possible to obtain a few general ideas on the processes involved. Thus it is fairly generally established that the specific conductivity of metal in the form of thin films is very much less than that of the same metal in bulk. The ratio of the two conductivities may be as large as 10^6; indeed the metal in the films has been described as a "non-conducting modification". This phenomenon has been mentioned in papers dating back some years. Within the last year or two it has been investigated in some detail on a variety of materials. J. Kramer[1] has obtained films of low conductivity by both sputtering and volatilization of the following metals: Fe, Ni, Pt, Zn, Sn, Cd and Sb. R. C. L. Bosworth[2] has observed low conductivity in nickel films prepared by electrodeposition on stainless steel. Devaux and Cayrel[3] report similar phenomena in thin films of CuS prepared by passing H_2S over the surface of a copper solution. Thus the phenomenon is not confined to films prepared in one particular way.

It is also fairly generally agreed that when the temperature of such films is raised, the conductivity increases, i.e. the temperature-coefficient of resistance is of the opposite sign to that of metal in bulk. Further, it frequently happens that when the temperature reaches a certain value, the conductance of the metal increases more or less suddenly to a comparatively high value, and on cooling this high value is retained. This phenomenon has been observed in varying degrees by Zahn and Kramer[6], Devaux and Cayrel[4], Wenk[5], Deubner[7], Deaglio[8] and Kramer[1]. According to Kramer the sudden large increase of conductivity occurs at a definite temperature which he calls the transformation-temperature T. This is considered to be characteristic of the metal, since Zahn and Kramer obtained the same value of T for films of Sb and Pt deposited electrolytically on Wollaston wire, as had been previously obtained for films of the same metals prepared by sputtering and by volatilization.

The most obvious suggestion as to the cause of the low conductivity of the films, when first made, is that the constituent particles are not in good contact. This was the view of some of the earlier workers, but it is not now common, if only for the reason that it may be observed in films as thick as 1μ. In order to investigate this point H. Murmann[9] has deduced the value of the electrical conductivity of thin

silver films from measurements of their transparency to heat radiation of long wavelength ($\lambda > 28\mu$). He found that the conductivity diminished very rapidly as the thickness decreased from 0·025 to 0·01 μ, and that it was of the order of 10^{3}–10^{4} times less than that of the metal in bulk. Since the conductivity determined in this way must be that of the constituent molecular groups of the film, no explanation based on contacts will hold.

The explanation advanced by Zahn and Kramer is that below the transformation-temperature metals may exist in an amorphous form which is non-conducting owing to the absence of free electrons. Kramer points out that metallic vapours are not conducting and considered that the condition of the metal in films of low conductivity is analogous to that of a highly compressed gas. As the temperature of the film increases the energy of the atoms increases until at the transformation-temperature T it reaches the value required for ionization: the conduction-electrons are released, and the crystal lattice formed. The conductivity then has its normal value and diminishes with further rise of temperature. As a result of measurements on seven metals Kramer states that the normal conductivity σ and T may be connected by the empirical formula

$$\sigma = \frac{CN^{\frac{1}{3}}}{T\alpha^{\frac{1}{3}}},$$

where C is a constant, N an integer, 1, 2, or 3, and α the atomic volume. A comparison of this with Sommerfeld's formula suggests that N is the number of free electrons per atom, and that the mean free path of the electrons in the normal metal is inversely propotional to T, which is perhaps plausible. It would mean that when the force holding the electrons to the ions before the transformation is large, the effect of the ions on the motion of the electrons after the transformation is also large, and the mean free path correspondingly small.

In order to verify his hypotheses of an amorphous phase Kramer made an X-ray examination of films of Sb in the two conditions. A film in the condition of low conductivity showed no Sb lines, while a film which had been heated beyond the transformation-temperature showed the lines distinctly. The amorphous modifications of Fe, Ni and Sb were found to be sensitive to light, the resistance diminishing with illumination. After the transformation they were insensitive. An amorphous film of Fe also appeared to be abnormally sensitive to a transverse magnetic field.

Kramer's contention is that all metals are capable of existing in an amorphous, non-conducting form, which is obtained when a sample is formed from atoms whose energy is less than that corresponding to the transformation-temperature. Thus by keeping the temperature low he was able to form films of Fe, Ni, Pt, Zn, Sn, Cd and Sb of abnormally low conductivity. He could not obtain such films for Ag or Au, which is explained on his theory by the very low transformation-temperature of these metals (estimated to be in the region of liquid-air temperatures). The investigations with which he has been connected, taken alone, suggest that his theory is adequate. It can also claim some measure of support from the work of Devaux and Cayrel and Bosworth, but these and other investigations suggest that the theory

is far from covering all the observed facts. Thus Bosworth found that the *resistance* of a Ni film of thickness 0·08 μ rose rapidly to a very high value on heating to 50° C. This effect disappeared on annealing at 80–100° but reappeared after a few days. Films of other thicknesses did not show it. Again Deubner made silver films by evaporation. Unlike Kramer he *did* find them to be of low conductivity and the conductivity increased about a thousandfold on heating to 70° for a few minutes. Longer heating to a higher temperature reduced the conductivity to a lower value than the initial one, i.e. as in Bosworth's experiment there was an increase of resistance. Gold behaved similarly but required higher temperatures. Again Devaux and Cayrel found that the high conductivity produced by heating their films of CuS was partly reduced by age, but restored by another heating. These experiments and also those of Perucca and other Italian workers discussed below suggest that the matter is much more complicated than Kramer's theory suggests. It should be pointed out, however, that Zahn and Kramer state that the characteristic transformation-temperature T is only applicable to films of a certain minimum thickness, since a certain minimum number of atoms will be necessary to ensure the formation of the ordinary lattice. The transformation-temperature may therefore be expected to vary with thickness for very thin films.

Another property which has received a good deal of attention is the variation of the properties of the film, when it is charged to a high voltage. Thus Perucca,[10] working with thin films of gold and platinum sputtered on to quartz-fibres, finds that the resistance of the film diminishes when it is charged to a potential of 4000 volts. The change of resistance is stated to be independent of the sign of the charge and to amount to as much as 40 per cent. Similar effects are reported by Deaglio[11] and Pierucci[12]. Deaglio[11] used films varying in resistance from 10^7 to 10^{14} ohms and he found that the effect was only large for films of resistance 10^{10}–10^{13} ohms, i.e. presumably for a certain range of thickness. The maximum change observed was about 40 per cent when the electrostatic field normal to the surface of the film amounted to 5000 volts per cm. Pierucci[12] claims that still larger changes are observed when the sign of the charge is varied several times. All these experiments were made on very thin films of gold, platinum or tungsten sputtered on to glass or quartz-fibres. Results of quite a different order are reported by A. Deubner[7] who worked with films deposited on a short length of glass tube by evaporation. The idea underlying his experiments was that when the metal film is positively charged, it is deprived of a certain number of free electrons and its conductivity may therefore be diminished. The measurements were made by means of an a.-c. bridge, with an amplifier and vibration galvanometer as detector. A silver film showed an increase of resistance of 1 part in 10^4 when it was charged to a potential of $+5000$ volts. A charge to a potential of -5000 volts gave a decrease of resistance of about the same amount, and the change was approximately linear with voltage. With gold and bismuth the effect was only about one-tenth of the above amount, while no change whatever could be detected with a copper film. The films of copper and bismuth were less stable than those of gold and platinum. These films were of considerably lower resistance (about 1000 ohms) than those of the Italian workers. The thickness,

estimated from the total weight of metal evaporated is stated to be of the order of 40 atoms. The measurements were made on the material in its normal form, i.e. after heating, although those of Cu and Bi showed no fall of resistance on heating.

Zahn and Kramer[6] have investigated the effect of charging films in another way. A flat film was supported close to a metal plate so as to form an air condenser. The capacity of this condenser was compared with that of a similar condenser formed from two metal plates, for a series of increasing applied voltages. At a certain voltage the capacity of the condenser of which the film was the high-potential electrode, began to decrease with increasing voltage. This is taken as an indication that no more free electrons were available for conduction from the film at this stage. The effect was detected with films of Bi and of Cu and as was expected, the critical voltage was lower the thinner the film. It is estimated that the Bi film must have contained 10^6 atoms to 1 free electron and that it must therefore be considered as amorphous.

Perucca[13] claims that Zahn and Kramer's hypothesis of an amorphous phase does not provide an adequate explanation of the facts. He regards the special properties of very thin films as merely those of all surface molecules and ascribes the many curious phenomena observed to orientation of the molecules, such as has been investigated by Finch, G. P. Thomson[14] and others, who have shown that films of platinum sputtered in various gases, when examined by electron diffraction, often show patterns indicating that they are composed of very small crystals (of the order of 5×10^{-7} cm.) which are oriented so that one face is parallel to the surface on to which they are sputtered. It is considered that the metals are sputtered as atoms, and form films composed of small crystals. The heating of the surface causes a growth of the crystals, and frequently results in orientation. The way in which these changes produce the various changes in electrical properties is by no means clear, but it is not improbable that they hold the key to the explanation.

Resistors of high value. The above work on thin metallic films has a practical bearing on electrical measurements of several types. For example, in the measurement of very small currents (ionization currents and the like) it is useful to have fixed resistances of very high values, e.g. 10^{10} or 10^{12} ohms. These are usually obtained by the deposition of thin films of conducting material on glass, porcelain or quartz in the form of short rods or fibres. There are evidently several dangers to guard against in the use of these resistors. The material used should be capable of giving a stable film: Deubner[7] found films of Cu to be less stable than those of silver and gold. He also found very thin films of platinum to be unstable, but Perucca[15] states that such films take a permanent value ($\pm 0{\cdot}5$ per cent after several days' ageing). The question of temperature-coefficient is also important: the value may be very much greater than that of the metal in bulk. Perucca[15] has found gold to be much better than platinum in this respect. L. F. Curtis[16] has found graphite very satisfactory for such resistors: graphite from a 3H drawing pencil is rubbed in longitudinal grooves in a pyrex rod, and is then sealed in glyptal lacquer. He mentions that Indian ink and lacquers other than glyptal were less satisfactory.

Resistors of the film type are also widely used in high-frequency measurements, in which it is of the utmost importance to reduce the capacitance and inductance of

the resistor to the smallest possible value, and also to avoid a change of resistance with frequency due to the "skin-effect". A thin film of short length and width satisfies these requirements, and at the same time will give values of resistance from say 100 ohms to many megohms. Graffunder[17] has described the construction of resistors of sputtered platinum for this purpose. Wenk and Wien[18] have used films of platinum-silver sputtered on to a ceramic material. Wenk[5] subsequently found that the resistors were unstable unless they were heated after sputtering. The general trend of the work on thin films suggests that such heating would be desirable in all cases, where it is desired to obtain a stable film of low temperature-coefficient; also that the thinnest films should be avoided where possible.

§ 5. RESISTANCE STANDARDS

The discovery of manganin as a material for resistance standards more than forty years ago led to a big improvement in the accuracy of measurement of electrical resistance, for this material had a very low temperature coefficient of resistance and also a small thermo-e.m.f. against copper. The standard coils made by Wolff of Berlin at that time have until quite recently not been improved upon, except in small points of detail (the use of potential terminals and the sealing of the coils in an oil-filled container as a protection against changes of atmospheric humidity). These coils are however found to show small secular changes,[1,2] and with the increasing demand for higher accuracy in recent years, several attempts have been made to produce more stable standards. The wire of the old coils was insulated with silk, and after winding, the coil was heavily coated with shellac varnish and then baked for some time at about 150° C. so as to anneal it as far as is possible without damaging the silk insulation. Obviously the annealing could only be very imperfect. In 1930 J. L. Thomas[3] devised a form of construction which allowed the wire to be thoroughly annealed after the winding of the coil, so that the strain produced in winding was completely relieved. The coil was of bare wire, which was wound on a mandrel and annealed at 550° C. in a vacuum furnace. It was then slipped on to its permanent insulated support, and hermetically sealed in an enclosure containing air only. Coils so made were considerably better than those of the older type, except when the wire was thin. It seemed probable that the thermal treatment required to produce a stable coil varied with the size of the wire. Experiments on similar lines are in hand at the N.P.L.[4] In 1934 Thomas[5] has described an even more promising development. Following recent work of Linde[6] on the electrical properties of various gold alloys, he has investigated in detail gold alloys containing 1·6–2·5 per cent of chromium, and finds that an alloy can be obtained with a temperature coefficient very much smaller than that of manganin. Information as to its stability can only be acquired after years of observation, but the results so far obtained suggest that it will probably be better than manganin in this respect also. The thermo-e.m.f. against copper is unfortunately three or four times as large as that of manganin, but this could be tolerated in a reference standard. The resistance R_t at any temperature t in the working range, 20–30° C. may be expressed in the form

$$R_t = R_{25}\,[1+\alpha\,(t-25)+\beta\,(t-25)^2].$$

The value of β is zero for 2·1 per cent chromium. The value of α for this alloy is also very small and may be reduced to zero by baking at say 150° C. for an appropriate time. In this way coils have been obtained with a resistance which is independent of temperature to a few parts in 10^7 over the range 20 to 30° C.

The work on the old resistance standards of the British Association[1] suggested that resistance coils of pure metal (particularly platinum) would be more stable than alloy coils. New standard coils of pure platinum have been constructed at the N.P.L.[7] while Thomas[8] at the National Bureau of Standards, Washington, has investigated standards of copper, gold, silver, tin and platinum. Of these, the silver and tin were definitely unsatisfactory; the gold and platinum gave promising results, the platinum being the better. Such coils would, of course, be useless for general purposes, owing to their large temperature-coefficient. It is necessary to use them at the temperature of melting ice. Thomas[9] estimates that the ice-point can be reproduced to almost 0·0001° C. using selected commercial ice wetted with distilled water, provided that a correction is made for barometric pressure. Dissolved air affects the temperature appreciably but in practice the mixture is always air-saturated.

H. v. Steinwehr[10] has given some useful data concerning the properties of standard resistors required to carry large currents for long or short periods. F. Wenner and J. L. Thomas[11] have pointed out that it is possible to adjust the resistance values or temperature-coefficients of four-terminal resistors, consisting of two or more resistance elements of manganin connected in parallel by copper bars or posts, by changing the position of the points of contact of the leads to the potential terminals and the copper posts, or reducing the cross-section of these posts. The temperature-coefficient may be either increased or diminished and may therefore often be reduced to zero.

§6. NEW INSULATING MATERIALS

In the last three or four years several new insulating materials have been produced which may have important applications in the construction of electrical apparatus such as electrometers, standard condensers, and all apparatus for use at high frequencies. For d.-c. work, amber has usually been regarded as the best insulator, although its high cost has limited its use to the more expensive standard instruments. Amber is a natural resin: the well known synthetic resins obtained as condensation products from phenol and formaldehyde (bakelite, etc.) while satisfactory as substitutes for amber for many purposes, have never been made with a resistivity as high as that of amber, so that they have not been suitable for use in instruments in which perfect insulation is of the first importance. Another series of resins, styrol[1, 2] resins, which has recently been produced, is decidedly promising in this respect. They are made by the polymerization of styrol or vinyl benzene $C_6H_5.CH : CH_2$, so that they are pure hydrocarbons, and as such may be expected to have excellent insulating properties; for the liquid hydrocarbons, benzene, hexane, transformer oil, etc., possess conductivities lower than those of other

liquids, while paraffin-wax is very good among the solids, although as ordinarily supplied, it does not appear to be stable, and has poor mechanical properties. The styrol resins are obtainable as clear glass-like solids. Perhaps the best known for electrical work is one of German manufacture sold under the trade name of "trolitul". It is sold as a powder, which softens on heating and can be moulded to shape. The mouldings are easily machined. Its specific gravity is given as 1·05, breakdown voltage 50 kV./mm.; dielectric constant 2·2 to 2·5 and the resistivity is extremely high, although it is not possible to quote a value. If it should prove to be permanent under ordinary laboratory conditions it will undoubtedly prove to be of great value. A similar material is manufactured in America under the name of "victron".

Considerable developments have also taken place in the manufacture of ceramic insulating materials[3,4,5,6] mainly intended for use at very high frequencies. Here d.-c. resistivity is not of very great importance. The effective conductivity when the voltage is alternating is proportional to the product of the dielectric constant, the power factor, and the frequency; and the quality of the material at any frequency is best judged by the product of the dielectric constant and power factor at that frequency. For such work fused quartz is the best-known material, although it is inferior to good amber for d.-c. work, probably because its surface tends to be more conducting than that of amber, owing to the formation of surface films of moisture. The new materials, some of which appear to be almost as good as fused quartz, consist chiefly of magnesium silicate, obtained from steatite or soapstone and talc. They are shaped by the methods of the ceramic industry, i.e. turning, casting, extruding and pressing, and then fired. After firing they cannot be readily worked, although they can be ground. They are sold in various grades, under trade names such as calan, and frequenta. Those made from talc (e.g. calan) are of a pure white colour, and appear to be rather better than those made from soapstone or steatite (e.g. frequenta) which are usually of a yellowish tint. All these materials are of course very suitable for work at high temperatures.

Another series of ceramic materials[3,4,6] of great interest are those containing titanium oxide as a base. These possess very high values of dielectric constant—values between 20 and 100 may be obtained—and at the same time the power factor is not unduly high; indeed it is considerably lower than those of the commoner insulating materials, such as ebonite, bakelite, etc. These materials are used chiefly for the making of condensers of small bulk. In this connexion it should be noted that their temperature-coefficient of dielectric constant is of the order -7×10^{-4} per 1° C. The corresponding value for mica is much smaller than this and of the opposite sign. The negative sign suggests that the material might be useful for the compensation of the temperature-coefficients of various materials such as standard air condensers.

A useful property of the magnesium silicate materials is that certain metals such as silver may be fused on to their surfaces. Thus condensers may be made by fusing metal electrodes on to the surfaces of a plate of the material. Coils of fixed inductance may also be made similarly, i.e. by fusing the metal turns of the coil on to cylinders,

discs, or rings of the ceramic material[2,6]. As the temperature-coefficient of expansion of the material is small ($2{\cdot}5$ to 9×10^{-8}) and the mechanical properties are very good, this process has some advantages for the construction of standard coils.

These materials may also be fused to certain kinds of glass, and the joints are gas-tight[5]. It thus becomes possible to seal electrodes into tubes of ceramic material by fusing ordinary glass-metal seals on to the tube. Alternatively, electrodes may be made by fusing a thin coating of silver on to rods of the ceramic material, and these electrodes may be sealed into the vessel, using glass as intermediate material. An electrode of this form is strong mechanically, and very suitable for the carrying of currents of very high frequency. For work involving such currents the process may have many applications.

It may be useful to tabulate some of the values given for the properties of the new materials, with those of a few other materials for comparison. It is probable that some of the values for the new materials only hold good for specially selected specimens. It is also very doubtful whether the accuracy implied by the values can be obtained, but even allowing a considerable margin for such factors, the properties of the materials must be considered remarkably good.

Material	Dielectric constant	Power factor: tan $\delta \times 10^4$ at frequency:			Reference
		10^6 c./sec.	10^7 c./sec.	10^8 c./sec.	
Quartz	4·7	1·1	1·1	1·0	4
Mica	7·0	1·7	1·7	1·6	4
Ultra-calan	7·1	1·1	1·1	1·0	4
Calan	6·6	3·2	2·6	—	6
Frequenta	5·6	4·2	3·4	3·0	4
Trolitul	2·2	4·1	3·7	—	6
Mycalex	8	19	18	18	4
Ebonite	3	64	107	150	4
Bakelite	2·8	160	220	500	4
Condensa	90	6	3·2	2·8	4

§ 7. MAGNETIC MEASUREMENTS

The measurement of magnetic field distortion. Prof. Rankine, in his Presidential address[1] to the Physical Society, discussed the possibilities of the measurement of irregularities of the earth's magnetic field by means of a magnetic torsion balance analogous to the balances so successfully used for gravitational work. The magnetic analogue to the Cavendish balance would consist ideally of two equal magnetic poles of the same kind, situated near the ends of a horizontal beam suspended by a torsion fibre. This might be realized by forming the beam of two equal coaxial magnets, with, say, their south poles adjacent and near the axis of suspension, so that they would have very little effect, their north poles near the ends forming the active portions. Such a balance would experience no resultant turning moment in a uniform field, but certain kinds of irregularities of the field would give rise to a differential couple tending to twist the suspension just as in the gravitational experiments. Very great sensitivity could be obtained with modern magnet steels, but the fact that the earth's

magnetic field is not in general vertical introduces very considerable practical difficulties; for unless the resultant moment of the beam were reduced to something of the order of one-millionth of the moment of the two constituent magnets, the turning moment due to the horizontal component of the earth's field would be likely to be large compared with that due to the irregularities of the field, so that the measurement of the latter would be impossible. Prof. Rankine concludes that on the whole there is little prospect of constructing a magnetic balance of this form with the necessary precision.

There is, however, a much better prospect of constructing a magnetic balance working on the Eötvös principle, which would measure the gradient $\partial H/\partial z$ or $\partial V/\partial s$ where H is the horizontal component and V the vertical component of the earth's field, z denotes the vertical direction, and s that of maximum variation of H. The magnetic analogue of the Eötvös balance designed by Prof. Rankine consists of a horizontal beam of non-ferromagnetic material suspended by a torsion wire or strip from a point vertically above its centre of gravity. From one end of the beam is suspended, also by a fibre, a permanent magnet with its axis vertical, or nearly so: actually there will always be a slight tilt of this magnet due to H, apart from the practical difficulties of mounting. The other end of the beam carries a counterpoise of non-magnetic material. The two equal and opposite poles of the magnet at the one end of the beam are clearly equivalent to two equal and similar poles at opposite ends and as these poles are at different levels, this simple system is equivalent to the Eötvös balance. Unfortunately, any slight departure of the magnet from the vertical gives it a horizontal component of moment, resulting in a twist of its suspension due to H, and therefore undesired torsion of the central suspension. It has not yet been found possible to satisfy the required conditions sufficiently closely to make the instrument suitable for magnetic surveying, but even with its present imperfections it has proved to be an admirable laboratory instrument[2] for measuring magnetic field distortion deliberately caused in the neighbourhood of the apparatus, and so sensitive that it may be used with the greatest ease for the detection of paramagnetism or diamagnetism in bodies such as water, glass, etc. The instrument takes up a definite orientation, determined by the residual horizontal component of moment and H, but is remarkably stable to all large-scale fluctuations of field, such as diurnal variations, magnetic storms, or those due to electric trains. This of course arises from the fact that the two poles of the magnet are exactly equal and that a field which is uniform over the space occupied by the instrument exerts no translational force on the magnet, and therefore no force tending to turn the beam. Movements of iron bodies in the same building cause a considerable deflection, and even bodies of very small susceptibility give a measurable deflection if they are placed quite close to the magnet. The instrument clearly possesses many of the advantages of that due to Kapitza and Webster, and described in last year's *Report*. It also makes possible the measurement of very small susceptibilities with weak fields. The possible use of the instrument as a very sensitive galvanometer of long period has also been suggested.

The measurement of the susceptibilities of liquids. G. Dupouy and C. Haenny[3]

have described a method for the measurement of the susceptibilities of liquids, based on that due to Gouy. A bar of quartz $18 \times 0{\cdot}5 \times 0{\cdot}5$ cm. is suspended in the field of a large electromagnet by means of a platinum wire attached to one pan of a balance. The axis of the bar is vertical and the upper end is arranged to lie in the central uniform part of the field of intensity H. The length of the bar is such that the lower end is outside the intense field of the electromagnet but lies in a comparatively weak field of intensity h. If S is the cross-section of the bar, then under these conditions it is subjected to an axial force f given by

$$f = \tfrac{1}{2}(\chi_s - \chi_0)(H^2 - h^2) S$$

where χ_s and χ_0 are the susceptibilities of the bar and the medium in which it is suspended respectively. This force may be balanced by means of weights on one of the pans of the balance and so measured. The liquid to be measured is placed in a glass tube fixed in the gap between the poles of the magnet and surrounding the bar. The force measured when the bar is suspended in the liquid gives $\chi_s - \chi_0$, and χ_s for the quartz being known, χ_0 for the liquid is determined. It is of course essential that the bar be homogeneous and of uniform cross-section. It was therefore cut from a good crystal, and the sides optically worked. With the powerful field available a mass susceptibility of 10^{-5} gave a force of 1 gm. and a change of mass susceptibility of 10^{-9} could be detected. The method is therefore suitable for work of high precision. The liquid was surrounded by an oil-bath by means of which it could be brought to any desired temperature, and the variation of magnetic properties with temperature could be studied.

§ 8. MEASUREMENTS ON FERROMAGNETIC MATERIALS

The methods for the measurement of the magnetic properties of ferromagnetic materials under ordinary conditions have now become more or less standardized, and developments of a fundamental character are not to be expected. Thus for ordinary measurements of permeability the well-known ballistic and magnetometer methods are available; while for measurements in very weak fields, null-methods involving the use of alternating current are employed, owing to the ease with which minute alternating currents may be detected by means of a telephone or vibration galvanometer. Recent developments consist largely of special arrangements of these standard methods, devised in order to cope with new or unusual conditions. Several developments of this character have been described during 1934. Thus W. B. Ellwood[1] has shown how the ballistic method may be used for flux-densities down to 0·0002 gauss. The novel feature is a ballistic galvanometer working on the resonance principle in much the same way as does the ordinary vibration galvanometer. Indeed the new galvanometer is essentially a vibration galvanometer designed to work at very low frequencies. The galvanometer is connected to a search coil in the ordinary way and the magnetizing current is reversed at regular intervals, equal to half the period of the galvanometer. Thus the moving coil receives an impulse every time it swings through the zero position, and the deflection therefore steadily increases until a final value, limited only by the damping of the moving system, is

reached. The necessary reversals may be made by hand, or by means of a photo-electric cell and relay controlled by the galvanometer deflection. In order to reduce the damping of the galvanometer to the lowest possible value, the suspension is mounted in an evacuated glass tube, so that air-damping is negligible. The magnetic field of the galvanometer is provided by a pair of Helmholtz coils carrying current instead of the ordinary permanent magnet. Thus all hysteresis effects in the magnet system of the galvanometer are avoided, and it therefore integrates correctly, and may be used without error for null-methods.

McKeehan[2] has described a magnetometer method suitable for measurements on very small specimens. He finds that large mono-crystals are not magnetically homogeneous and that it is therefore desirable for work on such materials to use specimens with linear dimensions not exceeding a few millimetres. For this purpose he uses a modification of the pendulum magnetometer of Weiss and Foëx. A light wooden bar is suspended in a horizontal position by five silk fibres—a pair forming a V at each end, with an additional one attached to one end in such a manner as to prevent rotation. The arrangement constitutes a pendulum, the deflection of which is observed by means of a microscope. The small specimens, in the form of oblate spheroids, are mounted on the bar, which passes through a system of coils arranged to produce a known magnetic field in the region occupied by the specimen. The force on the specimen in the direction z of the length of the bar may be represented by the usual formula

$$F_z = K \int_v H \frac{\partial H}{\partial z} \, dv,$$

which becomes

$$I_p \int_v \frac{\partial H}{\partial z} \, dv$$

if I_p, the component of the intensity of magnetization in the z direction, is independent of z, and H varies by an extremely small fraction throughout the volume of the test piece. This force is measured by the deflection of the pendulum. The coil-system consists of a solenoid, which controls the value of H, and a pair of Helmholtz coils in opposition which controls the value of $\partial H/\partial z$. These two quantities being separately controlled, it is easy to obtain a convenient deflection for a wide range of values of I_p. Temperature effects may be investigated by placing the specimen in a spoon of fused silica mounted at one end of the bar in a manner convenient for insertion into a small furnace.

A device for making measurements of the variation of permeability with tension has been described by Kelsall and Williams[3]. It is used for material in the form of thin strip or tape, which is bent into a triangle, the overlapping ends being spot-welded, so that a closed magnetic circuit is formed. Tension is applied at the corners in directions bisecting the angles. The magnetic measurements may be made by any standard method applicable to ring specimens, the triangle being treated as a ring. The use of a closed magnetic circuit for these measurements instead of the long straight strips often used for tension investigations has the

advantage that the material does not become magnetized by the earth's field. The more compact form is also convenient for the annealing of the material.

Changes of magnetic properties with time. In discussions of the magnetization of materials it is usually not necessary to consider time as a significant variable. The application of a magnetizing force H to a material may be regarded as producing instantly a definite magnetic induction B, and any change ΔH produces instantly a corresponding change ΔB. Thus the permeability $\mu = B/H$ and the incremental permeability $\Delta B/\Delta H$ may usually be considered as independent of time. It is known, however, that the time factor enters the problem in various ways.

(*a*) The magnetic induction B produced by a given H may increase comparatively slowly with the duration of application of H, reaching a final constant value only after long application of the magnetizing force. This phenomenon is often termed "magnetic after-effect" and is ascribed to "magnetic viscosity". It is seldom conspicuous, and when changes of B are observed, it is often difficult to make sure that they are not due to corresponding changes of H arising from the decay of eddy currents induced in the apparatus or specimen. However, Mitkevitch[4] claims to have observed changes due to magnetic viscosity in thin iron wire (0·1 mm. diameter) for periods up to 30 minutes after changes of H in the region 5 to 80 oersteds. When a material possessing these properties is subjected to cyclic changes of magnetizing force, the induction B must lag behind H by an amount depending on the period of the cycle, i.e. on the frequency, and there is therefore an energy loss per cycle, which is quite independent of ordinary magnetic hysteresis, and of eddy currents, and which varies with the frequency. It follows that the phenomenon may be conveniently investigated by a.c. methods. Investigations of this kind are discussed by Wittke[5] and Neumann[6].

(*b*) The permeability of certain ferromagnetic materials at low flux-densities is found to diminish with time after they have been subjected to the action of a comparatively large magnetizing force, or to severe mechanical treatment. It is as though the material were thrown into an unstable condition by the magnetic or mechanical disturbance, and gradually returns to the equilibrium condition when the disturbance ceases. The usual process of demagnetization of a specimen constitutes such a disturbance, and it is therefore necessary to note the time that has elapsed since the preliminary demagnetization, when making permeability measurements at low flux-densities. The effect is most noticeable in alloys containing a high percentage of silicon, and for values of B below 1000. Values of permeability more than twice the final equilibrium value have been noted. The excess of the permeability over its final value diminishes exponentially with time measured from the instant at which the disturbance ceases. The time-changes of permeability become smaller at high temperatures probably because the thermal agitation associated with a high temperature acts in the same way as a magnetic or mechanical disturbance. Investigations of these effects have recently been described by Webb and Ford[7] and by Sanford[8].

(*c*) Certain materials (soft irons) have long been described as showing "ageing" effects, viz. a diminution of permeability with time and an increase of hysteresis.

The question arises as to whether these changes are essentially the same as those of (*b*), the disturbance in this case being the annealing of the material. Wild and Perrier[9] regarded them as identical, since they could express the changes by the same formula. The information available as to ageing is, however, somewhat meagre. Webb and Ford regard the processes as different, although they do find that a high temperature may be regarded as a disturbing factor for process (*b*). There certainly does not appear to be any evidence suggesting that ageing can be overcome by magnetic treatment, as an alternative to annealing.

REFERENCES

§ 1. Measurement of voltage

(1) Perucca, E. *Z. InstrumKde* **47**, 524 (1927).
(2) Carmichael, H. *Proc. phys. Soc.* **44**, 400 (1932).
(3) —— *Proc. phys. Soc.* **46**, 169 (1934).
(4) Eggers, W. *Ann. Phys.* Lpz., **7**, 833 (1930).
(5) Wulf, T. Die Faden Elektrometer (1933). Berlin: Ferd. Dummlers.
(6) Nissen, H. F. *Hochfrequenztech. u. Elektroakust.* **42**, 24 (1933).
(7) Ulbricht, G. *Hochfrequenztech. u. Elektroakust.* **42**, 21 (1933).
(8) Palm, A. *Z. tech. Phys.* **15**, 117 (1934).
(9) Täuber-Gretler, A. *Bull. schweiz. elektrotech. Ver.* **25**, 556 (1934).
(10) Sharpe, B. A. *J. sci. Instrum.* **11**, 284 (1934).
(11) Bestelmeyer, A. *Z. tech. Phys.* **15**, 661 (1934).
(12) Wilkinson, E. *J. sci. Instrum.* **8**, 350 (1931).
(13) Richardson, J. E. *J. sci. Instrum.* **11**, 82 (1934).
(14) Hoffmann, G. *Phys. Z.* **13**, 480, 1029 (1912); *Ann. Phys.* Lpz., **42**, 1196 (1913).
(15) Engel, K. and Pforte, W. S. *Phys. Z.* **32**, 81 (1931).
(16) Michel, C. *Z. Phys.* **87**, 706 (1934).
(17) Herrmann, A. *Z. Phys.* **87**, 718 (1934).
(18) Banner, E. H. W. *J. sci. Instrum.* **11**, 218 (1934).
(19) Sims, L. G. A. *J. sci. Instrum.* **10**, 344 (1933).
(20) Stoner, C. R. and Grisdale, G. L. *J. sci. Instrum.* **11**, 313 (1934).
(21) Starr, A. T. *Proc. phys. Soc.* **46**, 35 (1934).
(22) Kirkpatrick, P. and Miyake, I. *Rev. sci. Instrum.* **3**, 1 (1932).
(23) Kirkpatrick, P. *Rev. sci. Instrum.* **3**, 431 (1932).
(24) —— *Rev. sci. Instrum.* **5**, 33 (1934).

§ 2. Voltage stabilization

(1) Körös, L. and Seidelbach, R. *Arch. Elektrotech.* **26**, 539 (1932).
(2) —— —— *Hochfrequenztech. u. Elecktroakust.* **40**, 9 (1932).
(3) *Marconi Rev.* Nos. 44, 45 (1933).
(4) Lämmchen, K. *Hochfrequenztech. u. Elektroakust.* **42**, 119 (1933).
(5) Meyer, H. *Bull. schweiz. elektrotech. Ver.* **25**, 516 (1934).
(6) Campbell, A. and Childs, E. C. *The Measurement of Inductance, Capacitance and Frequency*, p. 38 (1935). London: Macmillan.
(7) Evans, R. D. *Rev. sci. Instrum.* **5**, 371 (1934).
(8) Street, J. C. and Johnson, T. H. *J. Franklin Inst.* **214**, 155 (1932).
(9) Richards. *Rev. sci. Instrum.* **4**, 479 (1933).

§ 3. Measurement of current

(1) Curtis, H. L. and Curtis, R. W. *Bur. Stand. J. Res.*, Wash., **12**, 665 (1934).
(2) Snow, C. *Bur. Stand. J. Res.*, Wash., **11**, 681 (1933).
(3) Steinwehr, H. von. *Z. InstrumKde*, **6**, 249 (1932).

(4) VINAL, G. W. *Bur. Stand. J. Res.*, Wash., **8**, 729 (1932).
(5) VIGOUREUX, P. *Coll. Res. nat. phys. Lab.*, Lond., **24**, 81 (1932).
(6) HILL, A. V. *J. sci. Instrum.* **11**, 309 (1934).
(7) DISTAD, M. and WILLIAMS, J. H. *Rev. sci. Instrum.* **5**, 289 (1934).
(8) DONZELOT, P. and DIVOUX, J. *J. Phys. Radium*, **5**, 357 (1934).
(9) HASCHÉ, E. *Ann. Phys.*, Lpz., **19**, 484 (1934).
(10) DUBAR, L. *Rev. gén. Élect.* **34**, 669 (1933).
(11) HILL, A. V. *J. sci. Instrum.* **8**, 262 (1931).
(12) —— *Nature*, Lond., **133**, 685 (1934).
(13) —— *Nature*, Lond., **134**, 289 (1934).
(14) —— *J. sci. Instrum.* **11**, 246 (1934).
(15) —— *J. sci. Instrum.* **11**, 281 (1934).
(16) JONES, R. V. *J. sci. Instrum.* **11**, 302 (1934).

§ 4. ELECTRICAL CONDUCTIVITY OF THIN METALLIC FILMS

(1) KRAMER, J. *Ann. Phys.*, Lpz., **19**, 37 (1934).
(2) BOSWORTH, R. C. L. *Trans. Faraday Soc.* **30**, 549 (1934).
(3) DEVAUX, H. and CAYREL, J. *C.R. Açad. Sci.*, Paris, **198**, 1339 (1934).
(4) —— —— *C.R. Acad. Sci.*, Paris, **199**, 912 (1934).
(5) WENK, P. *Phys. Z.* **35**, 653 (1934).
(6) ZAHN, H. and KRAMER, J. *Z. Phys.* **86**, 413 (1933).
(7) DEUBNER, A. *Ann. Phys.*, Lpz., **20**, 449 (1934).
(8) DEAGLIO, R. *Nuovo Cim.* **11**, 447 (1934).
(9) MURMANN, H. *Z. Phys.* **89**, 426 (1934).
(10) PERUCCA, E. *C.R. Acad. Sci.*, Paris, **198**, 456 (1934).
(11) DEAGLIO, R. *Nuovo Cim.* **11**, 288 (1934).
(12) PIERUCCI, M. *Nuovo Cim.* **11**, 277 (1934).
(13) PERUCCA, E. *Nuovo Cim.* **11**, 531 (1934).
(14) THOMSON, G. P., STUART, N. and MURISON, C. A. *Proc. phys. Soc.* **45**, 381 (1933).
(15) PERUCCA, E. *C.R. Acad. Sci.*, Paris, **189**, 527 (1929).
(16) CURTIS, L. F. *Rev. sci. Instrum.* **4**, 679 (1933).
(17) GRAFFUNDER, W. *Ann. Phys.*, Lpz., **19**, 7, 689 (1934).
(18) WENK, P. and WIEN, M. *Phys. Z.* **35**, 145 (1934).

§ 5. RESISTANCE STANDARDS

(1) GLAZEBROOK, R. T. and HARTSHORN, L. *Rep. Brit. Ass.*, p. 418 (1932).
(2) PERARD, A. and ROMANOWSKI, M. *C.R. Acad. Sci.*, Paris, **199**, 523 (1934).
(3) THOMAS, J. L. *Bur. Stand. J. Res.*, Wash., **5**, 295 (1930).
(4) *Rep. nat. phys. Lab.*, Lond., p. 72 (1934).
(5) THOMAS, J. L. *Bur. Stand. J. Res.*, Wash., **13**, 681 (1934).
(6) LINDE, J. O. *Ann. Phys.*, Lpz., **402**, 52 (1931); **407**, 219 (1932).
(7) *Rep. nat. phys. Lab.*, Lond., p. 60 (1934).
(8) THOMAS, J. L. *Bur. Stand. J. Res.*, Wash., **12**, 313 (1934).
(9) —— *Bur. Stand. J. Res.*, Wash., **12**, 323 (1934).
(10) STEINWEHR, H. VON. *Z. InstrumKde*, **54**, 48 (1934).
(11) WENNER, F. and THOMAS, J. L. *Bur. Stand. J. Res.*, Wash., **12**, 147 (1934).

§ 6. NEW INSULATING MATERIALS

(1) LIGHT, L. *Brit. Plast. mould. Prod. Tr.* **4**, 134 (1932).
(2) ROHDE, L. *Z. tech. Phys.* **14**, 480 (1933).
(3) HANDREK, H. *Hochfrequenztech. u. Elektroakust.* **43**, 73 (1934).
(4) —— *Z. tech. Phys.* **15**, 491 (1934).
(5) —— *Z. tech. Phys.* **15**, 494 (1934).
(6) —— *Z. Ver. dtsch. Ing.* **78**, 1441 (1934).

§ 7. Magnetic measurements

(1) Rankine, A. O. *Proc. phys. Soc.* **46**, 1 (1934).
(2) —— *Proc. phys. Soc.* **46**, 391 (1934).
(3) Dupouy, G. and Haenny, C. *C.R. Acad. Sci.*, Paris, **199**, 781 (1934).

§ 8. Measurements on ferromagnetic materials

(1) Ellwood, W. B. *Rev. sci. Instrum.* **5**, 300 (1934).
(2) McKeehan, L. W. *Rev. sci. Instrum.* **5**, 265 (1934).
(3) Kelsall, G. A. and Williams, H. J. *Rev. sci. Instrum.* **5**, 437 (1934).
(4) Mitkevitch, A. *C.R. Acad. Sci., U.S.S.R.* **1**, 534 (1934); **3**, 425 (1934).
(5) Wittke, H. *Ann. Phys.*, Lpz., **18**, 679 (1933); **20**, 106 (1934).
(6) Neumann, E. A. *Z. Phys.* **89**, 308 (1934).
(7) Webb, C. E. and Ford, L. H. *J. Instn elect. Engrs*, **75**, 787 (1934).
(8) Sanford, R. L. *Bur. Stand. J. Res.*, Wash., **13**, 371 (1934).
(9) Wild, G. and Perrier, A. *Arch. Sci. phys. nat.* **7**, 209 (1925).

THE CHARGE OF THE ELECTRON

By Professor H. R. ROBINSON, F.R.S.,
Queen Mary College (University of London)

Throughout this Report the electronic charge is quoted in e.s.u., the specific charge in e.m.u. gm.$^{-1}$ *In references to the years* 1929 *onwards, absolute, not international, electrical units are implied.*

§ 1. INTRODUCTION

In 1929 Prof. R. T. Birge[1] published an acutely critical and masterly survey of our knowledge of the fundamental physical constants. It was a very timely summary, and it undoubtedly—if we may borrow from the vocabulary of another trade—did much to promote a desirable "constants-consciousness" in the general body of physicists. This especially applies to two of the atomic constants, the charge e and the specific charge e/m_0 of the electron.

At the time of Birge's report, there were two distinct, and apparently equally well authenticated, values of e/m_0. Direct deflection methods gave $1{\cdot}769\times10^7$, while spectroscopic methods gave $1{\cdot}761\times10^7$ abs. e.m.u. gm.$^{-1}$ As the maximum error admitted in each case was only of order 1 part in 1000, the relatively large difference between the two values was more than a little disturbing. On the other hand, Millikan's value of e (only slightly modified in revisions of the calculations, made by Birge[1] and by Millikan[2] himself) was generally accepted as accurate to within 1 part in 1000. In fact it seemed very likely that the electronic charge lay somewhere between $4{\cdot}768$ and $4{\cdot}772\times10^{-10}$ e.s.u. It is significant of the authority attaching to Millikan's work that it was considered necessary to correct his values for the small difference between absolute and international electrical units and for a small change in the accepted value of the velocity of light.

The situation has oddly changed since 1929, the spectroscopic and deflection values of e/m_0 now being in excellent agreement at something very near to $1{\cdot}758\times10^7$. There are, however, now in the field two values of e which differ by more than 7 parts in 1000; they bear in fact to one another almost exactly the celebrated and possibly significant ratio 136/137[3]. The first of these is Millikan's, the second is the value deduced from absolute X-ray wave-lengths by a method which was only beginning to be fully exploited at the time of Birge's first paper.

This situation justifies the publication of a brief *Report*. In view of the excellence of the summaries given by Birge and others, there is no point in discussing here any but the most direct—or apparently most direct—and accurate experimental methods for evaluating e. In the front rank of these come only (1) methods involving the observation of charged clouds or individual droplets, and (2) the comparison of "crystal" and "ruled-grating" wave-lengths of X rays.

The first of these, in various forms, has been in regular use ever since the time of the pioneer experiments of Townsend[4, 5] and J. J. Thomson[6] in 1897 and 1898. The details of its development in the hands of Millikan are now very well known to physicists—in fact they constitute one of the major epics of modern physical research. The second, being as yet less widely known, will be dealt with here in somewhat greater detail.

Reduced to its elements, the method is this: the wave-length of a selected monochromatic X radiation is determined in absolute measure by diffraction at a ruled grating of known spacing. The same radiation is reflected from a suitable crystal face, and from the measured reflecting angle, the corrected Bragg reflection law being assumed, the grating space of the crystal may be deduced at once. Hence from the type of structure of the crystal (assumed known from detailed X-ray analysis) and its density we find the absolute mass of one molecule of the crystal substance. Dividing this into the molecular weight we get Avogadro's number, and this again, divided into the Faraday constant, gives e.

For comparing the experimental data in the form in which they frequently appear, it is convenient to think of this calculation as carried out in the reversed sense, which happens to correspond more closely to the historical order of approach to the problem of measuring X-ray wave-lengths. Thus, the grating space of a crystal is equal to $\sqrt[3]{(M/\gamma\rho N)} = \sqrt[3]{(Me/\gamma\rho F)}$, where M is the molecular weight of the crystal substance, ρ the density, N Avogadro's number, and F Faraday's constant (measured, of course, in the same units as e). γ is a factor characteristic of the crystal, which can be determined unambiguously by X-ray spectrometry without assuming any absolute lengths, and which has a particularly simple form for a cubic crystal like rock salt.

In this way, the work of the Braggs and Moseley in 1913 established the approximate value $d_{100} = 2{\cdot}814 \times 10^{-8}$ cm. for the spacing of the (100) cleavage planes of the rock salt crystal. This is usually known as Moseley's[7] value of the rock salt grating space. It is evident that e is the only factor in which there can be serious uncertainty, and that a small error in e, say x per cent, leads to an error of $\frac{1}{3}x$ per cent, in the same sense, in the values of d and of the X-ray wave-lengths derived therefrom.

Moseley's d is quoted to 1 unit in approximately 2800, but the last figure is not significant in any sense. It has no right to be there at all, and quite possibly its presence merely indicates that Moseley used four-figure tables.* On the other hand, X-ray wave-lengths may be *compared* by the crystal spectrometer to an accuracy which easily exceeds 1 part in 25,000, and which may even approach 1 part in 250,000. Siegbahn therefore suggested some years ago that the grating space of rock salt at 18° C. should be taken *arbitrarily* as $2{\cdot}81400 \times 10^{-8}$ cm., and that all X-ray wave-lengths should be given in terms of this assumed grating space.

* It is interesting to note that in his paper Moseley, before embarking on the calculation of d, quotes two quite irreconcilable figures—namely, $N = 6{\cdot}05 \times 10^{23}$, $e = 4{\cdot}89 \times 10^{-10}$ (Millikan's(8) 1911 value). It is easy to verify that in the calculation he used the value of N quoted, not that of e. It is thus largely a fortunate chance that Moseley's grating constant agrees so closely with the values deduced by later workers—e.g. the value $2{\cdot}81447 \times 10^{-8}$ cm. calculated by Compton, Beets and De Foe[9] from more accurate data in 1925.

This is equivalent to defining the X unit (nominally 10^{-11} cm.) as $d_{100}/2814$ exactly, where d_{100} is the true grating space of rock salt at 18° C.

Suppose now that a certain value of e—say e_1—has been assumed in the course of calculating d, and that all other errors are small compared with the error in e_1. If now it is found that the true value of the X unit is $(1+\delta)\times 10^{-11}$ cm., it is obvious from the above that the X-ray experiments indicate $(1+\delta)^3\times e_1$ as the true value of e. This is a form in which the X-ray results are often expressed, as will be seen from examples to be quoted later.

The fact that the greater part of accurate X-ray spectrometry has been performed with calcite, and not rock salt, as the crystal does not vitally concern us here, as the relative dimensions of the two crystals are very well known, and results obtained with the one may readily be expressed in terms of the other. Corresponding to the value already quoted for rock salt, Siegbahn[10] finds for the cleavage plane of calcite at 18° C., $d=3029{\cdot}04$ X units.* Both figures are assumed to refer to first-order reflections. The corresponding value of d_∞ for very high orders of reflection by calcite is 3029·45 X units. These values (corrected when necessary for temperature) have been very widely used by X-ray spectroscopists, but it should be noted that Bearden, some of whose results will be quoted later, generally employs his own value, $d_\infty=3{\cdot}02810$ A. at 18° C. This value is based[11] upon a very careful examination of the structure and density of calcite crystals, and upon the assumption that $e=4{\cdot}768\times 10^{-10}$ e.s.u.

§ 2. THE OIL-DROP VALUE

We may now with advantage consider the general run of the numerical results obtained in the experiments with charged drops and with X rays respectively. A full historical account would be out of place here; it is, however, worth recalling that as early as 1898 Townsend[5] estimated $e=5\times 10^{-10}$ e.s.u. approximately, from his experiments on the clouds formed in gases freshly liberated by electrolysis. Townsend's earlier work[4] had indicated $e=3\times 10^{-10}$ e.s.u. approximately as the charge per droplet, but it was recognized that this result was vitiated by the presence in the clouds of charges of both signs. In the later paper Townsend somewhat tentatively corrected his result to the larger figure; in view of our ignorance of the mechanism by which the droplets acquire their charges, this can scarcely rank as a very direct determination of the electronic charge, but at least Townsend made a remarkably accurate forecast of the true value of e.

The first direct measurements, those of Sir J. J. Thomson[6] with charged water drops, published in 1898 and 1899, gave $6{\cdot}5$–$6{\cdot}7\times 10^{-10}$—a figure which was later corrected to $3{\cdot}4\times 10^{-10}$ as a result of further experiments[12] carried out in 1901–2. This was supported by the work of H. A. Wilson[13], in which for the first time vertical electric fields were used to modify the motions of the droplets under

* Compton, Beets and De Foe[9] calculated in 1925 for the grating space of calcite, $d=3{\cdot}02904\times 10^{-8}$ cm. at 18° C., apparently in exact agreement with Siegbahn's value. This is, however, no more than a very remarkable coincidence, as the calculated value of d represents the value which should correspond to very high orders of reflection, or d_∞. Compton, Beets and De Foe's value for first order reflection, d_1, is $3{\cdot}02858\times 10^{-8}$ cm. at 18° C.—about 1 part in 6500 less than Siegbahn's.

examination. Wilson's results ranged from 2·0 to $4{\cdot}4 \times 10^{-10}$, with a mean value of $3{\cdot}1 \times 10^{-10}$. The spread of the results gives some indication of the experimental difficulties—particularly that of working with rapidly evaporating drops.

As a result of the work of Thomson and Wilson, a low value of e was generally accepted during the earlier years of this century. The first experiments of Millikan[14], made in collaboration with Begeman, led however to an appreciably higher value, $4{\cdot}03 \times 10^{-10}$. A little later (1910), Millikan[15] obtained $4{\cdot}65 \times 10^{-10}$, and at that time he was able to quote five recent and independent determinations, all lying between $4{\cdot}65$ and $4{\cdot}79 \times 10^{-10}$, with a mean value of $4{\cdot}69 \times 10^{-10}$.

This brings us to the most important stages of Millikan's work, in which he perfected the method of observation of individual droplets in very accurately standardized electric fields, and in which he used relatively stable drops of oil or mercury. These drops were kept under observation for long periods, during which the temperature and pressure in the surrounding air were very accurately controlled, in a manner which had been quite impossible in the earlier experiments with water drops. The coefficient of viscosity of air was specially redetermined, and the law of the resistance of air to the slow motion through it of small drops was very thoroughly examined. The work is excellently summarised in Millikan's well-known book[16].

As the result of this work, Millikan gave in 1913[17] the value

$$e = (4{\cdot}774 \pm 0{\cdot}009) \times 10^{-10} \text{ e.s.u.}$$

In 1917[18], after a careful revision of the work, he confirmed this value, with an even smaller margin of error,

$$e = (4{\cdot}774 \pm 0{\cdot}005) \times 10^{-10} \text{ e.s.u.}$$

This—often quoted as "the 1917 value"—has been slightly modified[1,2] by manipulating Millikan's original data in different ways, and by using more recent values for the electrical units and the speed of light. Apart from these minor changes, it still stands as the representative oil-drop value of e. Other workers have, it is true, conducted similar researches, either to measure e or to find the appropriate modification of Stokes' law to apply to very small spheres moving through a relatively coarse-grained medium. These other experiments were however conducted on a comparatively modest scale, and they do not approach Millikan's in final accuracy. They therefore do not contribute very directly to our knowledge of the precise limits between which e must lie, and there is no point in enumerating them here.

Millikan himself[2] upholds the change to $4{\cdot}770 \times 10^{-10}$ e.s.u. Birge[19], in an important paper published in 1932, gave $4{\cdot}7688 \times 10^{-10}$ as the most probable value. Another value suggested by Birge[20], namely $(4{\cdot}7668 \pm 0{\cdot}0038) \times 10^{-10}$, is generally considered to be too low.

To sum up this section of the *Report*: the value $\mathbf{4{\cdot}769 \times 10^{-10}}$ e.s.u. may fairly be taken as the oil-drop value of e. There is a possibility (*v. infra*) of an error in the coefficient of viscosity of air adopted in the calculations, but it does not seem likely that there can be any other errors totalling as much as 1 part in 1000.

§ 3. THE X-RAY VALUE

The accurate X-ray determinations of e are all fairly recent. The pioneer experiments of Compton and Doan[21] in 1925, which led the way into this new branch of X-ray spectroscopy, were just accurate enough to show the approximate equality (to within about $\pm 0{\cdot}5$ per cent) of crystal and ruled-grating wave-lengths. Wadlund's[22] experiments, which were among the first to be carried out under conditions specially designed for accuracy, led to results in almost exact agreement with the crystal values. In fact Wadlund's mean X-ray value of e was $(4{\cdot}774 \pm 0{\cdot}007) \times 10^{-10}$—identical with Millikan's, but with a slightly greater probable error.

The later work with ruled gratings has, I think without exception, yielded higher values of e. Bäcklin's[23] first experiments (1928) gave $e = 4{\cdot}793 \times 10^{-10}$ e.s.u.; Compton[24] (1929) gave $4{\cdot}810 \times 10^{-10}$, while Cork[25] obtained in 1930 the very high values $4{\cdot}821$ and $4{\cdot}816 \times 10^{-10}$. On the other hand, Shiba[26] has calculated $e = 4{\cdot}798 \times 10^{-10}$ from selected later results of Witmer and Cork[27]. In general, the presumably more accurate work of the last few years has produced results lying between $4{\cdot}800$ and $4{\cdot}810 \times 10^{-10}$. Bearden[28], who has since 1929 investigated many aspects of the problem of X-ray wave-lengths, obtained in 1931 the value $e = 4{\cdot}806 \times 10^{-10}$ e.s.u. This is very near the mean of the most recent and consistent values.

The newest results are:

Bäcklin[29]: 1 X unit $= 1{\cdot}00218 \times 10^{-11}$ cm., $e = 4{\cdot}805 \times 10^{-10}$;

Södermann[30]: 1 X unit $= 1{\cdot}00225 \times 10^{-11}$ cm., $e = 4{\cdot}806 \times 10^{-10}$;

Bearden[31]: $e = (4{\cdot}8036 \pm 0{\cdot}0005) \times 10^{-10}$.

Bäcklin's and Södermann's results are given as correct to 1 part in 8000 and 1 in 10,000 respectively for the X unit—say to 1 part in 3000 for e. Both are based on measurements of the aluminium Kα doublet—i.e. of a rather soft X radiation. Bearden's result is quoted in a short letter summarizing his recent work, which includes measurements with harder X rays.

To sum up, the X-ray determinations of e are remarkably concordant, over a quite wide range of wave-lengths. The mean of the best results lies very close to $\mathbf{4{\cdot}805 \times 10^{-10}}$ e.s.u.—probably accurate to within about 1 part in 4000.

§ 4. POSSIBLE SOURCES OF ERROR IN THE X-RAY AND OIL-DROP VALUES

The ratio of the oil-drop to the X-ray value is therefore very nearly $4{\cdot}769/4{\cdot}805 = 136/137{\cdot}0_3$ which is $136/137$ within the limits of experimental error. It has already been suggested[3] that the occurrence of this ratio may be significant. We shall not discuss this here, but shall simply examine the experimental evidence with a view to

seeing how far either value may admit of modification. It will be kept in mind that the difference between the two values is very nearly 0·75 per cent, corresponding to about 0·25 per cent in the X-ray wave-lengths.

(*a*) *The X-ray value*

No error even approaching this order of magnitude can be supposed to exist in the wave-length *measurements*, either with crystals or ruled gratings. The technique of X-ray crystal spectrometry is now very highly developed, and numerous examples of the accuracy easily attainable are to be found, for instance, in the standard treatises of Siegbahn[32] and of Compton and Allison[33]. The work with ruled gratings is in some respects more difficult than that with crystals, but there is convincing agreement among the recent results of different workers in this field. If there is any serious error in the X-ray result, it must originate in some defect of the theories used in working out the wave-lengths from the measurements. The chance of there being any serious defect seems, however, very small. The theory of the reflection of X rays by perfect crystals was put on a sound basis by Darwin[34], in two important papers which appeared at a very early stage in the history of X-ray spectroscopy. The theory has been further examined by a number of writers, and it has been shown that minor imperfections in crystal structure (mosaic and secondary structure), which have been studied theoretically by Zwicky and others, cannot appreciably affect the positions of the diffraction maxima.

Similarly, there is no serious source of error apparent in the theory of the ruled-grating measurements. Prof. A. W. Porter[35] first raised the interesting point that the ordinary theory of the diffraction grating, as used in normal optical practice with collimated beams, must be modified before it can be applied to the X-ray problem, where divergent pencils are necessarily used. The correction for divergence is, however, negligible in all the experiments quoted above. Further, possible errors due to ruling irregularities in the grating have been successfully eliminated.

There remains however one very plausible objection to the crystal wave-lengths. It may reasonably be assumed that in a crystal—as in any other body—the surface layers may differ appreciably from the deeper-lying portions. The grating constant d which is computed from the average density of the whole crystal may then be different from the effective mean spacing d_s of a set of planes of atoms near the surface. Since in most cases of X-ray reflection only a thin surface layer of the crystal is used, this assumption leads at once to a qualitative explanation of the discrepancy between crystal and grating wave-lengths. The discrepancy is removed entirely if we can assume that d_s differs from d by only 0·25 per cent.

Tu's experiments[36] show, however, that if we accept this hypothesis we must also assume that the ratio d_s/d is very accurately the same for four different crystals —viz. calcite, rock salt, sylvine and diamond—which at the least would be a most remarkable coincidence. Again, we should expect the properties of the crystal, and in particular d_s, to vary somewhat rapidly with the depth in this hypothetical transition layer. It turns out, however, that the wave-lengths measured in high-order reflections agree very closely with those measured in the first order (apart

from the relatively insignificant difference, about 1 part in 10,000, which is satisfactorily explained in terms of the refractive index of the crystal). As the different order reflections correspond to very different penetrations of the X rays into the crystal, this seems very definite evidence against the hypothesis which we have been momentarily (and somewhat inhospitably) entertaining.

We must conclude that at present the X-ray value of e appears to be unassailable alike on experimental and theoretical grounds.

(*b*) *The oil-drop value*

The oil-drop value has been attacked on two grounds. In 1929 H. A. Wilson[37] pointed out that some of Millikan's observations on the behaviour of drops with different values of pa (gas pressure × radius of drop) could very well be fitted to an empirical equation differing from that used by Millikan. Using the alternative equation with these observations, Wilson found $e = 4{\cdot}82 \times 10^{-10}$ e.s.u. This argument has however been met by Millikan[2].

Shiba[26] who, after sifting most of the evidence available in 1932–3, decided in favour of the X-ray value of e, concluded that Millikan took too low a value for the coefficient of viscosity (η) of air. In the oil-drop method, $\eta^{\frac{3}{2}}$ appears as a factor in the final expression for e; an increase of 0·5 per cent in the value adopted for η would therefore bring the oil-drop value of e into exact agreement with the X-ray value.

Millikan's 1917 result is based on the value of η obtained by Harrington[38] in 1916. This value was substantially confirmed in 1923 by Van Dyke[39], who used the same method (rotating cylinders). It is not difficult to find quoted values of η, obtained by other methods, which are higher than Harrington's by as much as 1 per cent. It is therefore possible that the discrepancy between the oil-drop and the X-ray values of e may ultimately be reduced as a result of improvements in our knowledge of the resistance offered by air to the slow passage of small spheres.

§ 5. OTHER METHODS OF ESTIMATING e

(*a*) *Measurement of charge of α particles*

It is singularly unfortunate, given the magnitude of the discrepancy, that outside these two methods there is available so little direct experimental evidence. Undoubtedly the most direct of all methods of finding e is that of counting the rate of emission of α particles from a suitable radioactive source, and measuring the total current carried by these α particles. It will be recalled that the first experiments of this kind, by Rutherford and Geiger[40] and by Regener[41], gave remarkably good values of e, though they were not designed for high precision. Most of the later work is summarized in the treatise of Rutherford, Chadwick and Ellis[42]. We need only mention here the measurements of current by Braddick and Cave[43], and of number of α particles by Ward, Wynn-Williams and Cave[44]. Both sets of experiments were conducted with every precaution to ensure accuracy, and (a most important point), as both were performed in the same laboratory, the resulting value of e is independent of the absolute value of the radium standard employed.

Braddick and Cave[43] deduced from their measurement of current that the number of α particles emitted per sec. per gm. of radium is $Q = 3{\cdot}68_2 \times 10^{10}$; in a footnote to [44] they correct this to $3{\cdot}69_0 \times 10^{10}$. A correction of 0·3 per cent was added to the measured values of the α-particle currents, to allow for the presence of a small proportion (estimated at 1 in 170) of singly charged helium atoms among the normal α particles. The value given for Q was based on the assumption that the charge of the normal α particle is $2 \times 4{\cdot}774 \times 10^{-10}$ e.s.u.

The direct counting experiments[44] gave $Q = 3{\cdot}66_3 \times 10^{10}$. Hence from these experiments we may deduce: $e = \frac{3{\cdot}69_0}{3{\cdot}66_3} \times 4{\cdot}774 \times 10^{-10} = 4{\cdot}809 \times 10^{-10}$ e.s.u.

This is in good agreement with the X-ray value, but it must be pointed out that the authors' estimate of the probable error in the counting experiments is rather large. The experiments are difficult, and liable to be vitiated by a number of parasitic effects. Further, it is necessary to count a very large number of α particles in order to obtain a representative value for the average rate of emission. In the experiments under discussion, in which 92,000 particles were counted, the probable error due to fluctuations in emission alone is 0·3 per cent. The authors estimate their final probable error in Q as ±0·5 per cent (apart from the possible inaccuracy of the radium standard, which does not concern us). There is also an appreciable error, of about the same order of magnitude, in Braddick and Cave's results, and it is at least possible that the agreement between the α-particle and the X-ray values of e is fortuitous.

Very recently Schopper[45] has redetermined the charge of the α particle. In December, 1934, he published the value $(4{\cdot}768 \pm 0{\cdot}005) \times 10^{-10}$ e.s.u. for the half-charge of the normal α particle—in excellent agreement with the oil-drop e. His paper has been criticized by Birge and McMillan[46], and in addition it appears from a later paper[47] that there was a serious and quite unsuspected source of error in the work; the standard cadmium cell used in the experiments is now found to be exhibiting quite abnormal and somewhat rapid variations in electromotive force.

So far, then, the most reliable evidence given by the α-particle measurements is in favour of the X-ray value of e. Schopper is repeating his experiments, and the results of his new work will be received with considerable interest.

(*b*) *Measurement of refractive index in the X-ray region*

Interesting, but not quite conclusive, evidence is provided by accurate measurements of the refractive indices of different materials for X rays. The X-ray wave-lengths deduced from Bearden's[48] first experiments with quartz were in very fair agreement with the crystal wave-lengths—and therefore with the oil-drop value of e. The more recent work of Bearden and Shaw[49], in which revised values are used for the constants appearing in the calculations, yields the result: 1 X unit $= 1{\cdot}0023 \times 10^{-11}$ cm.—in very close agreement with the grating measurements. Still more recently Bearden[50] has confirmed this by independent measurements of the refractive index of diamond for Cu $K\beta_1$ rays.

The measurements of refractive indices for X rays are clearly accurate, but it is not certain that the translation into wave-lengths is equally reliable. The existing theory of dispersion contains a number of debatable approximations, and further, in the numerical calculation of λ from the refractive index it is necessary to assume a value for $\sqrt{(e/m_0)}$. The probable error in the latter factor is small, but the element of uncertainty in the theory constitutes a more serious objection to this apparent confirmation of the X-ray value of e. It is clear, however, that Bearden's new results with the diamond prism do much to diminish the weight of this objection.

(*c*) *Energies of X-ray electrons*

Finally, we may mention some experiments on the energies of the photoelectrons ejected from elements by monochromatic X rays[51]. The energies are deduced from measurements of the deflections experienced by the electrons in known magnetic fields; they may then be compared with the energies related to known X-ray spectral terms. In making this comparison it is (on account of the way in which the photoelectron energies are measured) necessary to use a conversion factor which contains $e/h \times e/m_0$, if the X-ray terms are known absolutely; when the terms are deduced from crystal wave-lengths, there is an additional factor $e^{\frac{1}{2}}$. In effect, therefore, the photoelectric measurements, combined with the Moseley-Siegbahn wave-lengths, provide a means of determining the value of $(e^{\frac{1}{2}}/h) \times (e/m_0)$.

It is fairly well known that these experimental results[51] are difficult to reconcile with generally accepted values of the constants—the photoelectron energies being apparently too large. It is not easy to find the best way of using the results, as the quantity actually deduced from the experiments is a little too replete with constants—or combinations of constants, like e/h—with which one may play more or less at will. Ruark[52] has, however, recently indicated what is probably by far the best way of handling the experimental data. He starts with the X-ray value of e (which carries with it the grating wave-lengths), and obtains his conversion factor from the equation:

$$\frac{e^{\frac{1}{2}}}{h} = \left(\frac{Rc^2}{2\pi^2}\right)^{\frac{1}{2}} \left(\frac{e}{m_0}\right)^{\frac{1}{2}}.$$

He uses the Shane and Spedding[53] value of e/m_0 ($1{\cdot}7579 \times 10^7$ e.m.u. gm^{-1}), and the error in the resulting value adopted for the right-hand side of the equation is probably very small indeed. Ruark shows that with these values of the constants the photoelectric results are in excellent accord with the X-ray data. The corresponding value of e/h is however difficult to reconcile with other experimental results.

§ 6. SUMMARY

(1) The oil-drop and X-ray values of e are very nearly $4{\cdot}769$ and $4{\cdot}805 \times 10^{-10}$ e.s.u. respectively.

(2) The ratio of these numbers is almost exactly 136/137, which may be significant.

(3) The oil-drop value may be in error because of our insufficiently accurate knowledge of the viscous resistance experienced by small drops as they move through air.

(4) The X-ray value may be in error because of some unknown defect in the theory of the diffraction of X rays by ruled gratings or (less probably) in the theory of diffraction by crystals.

(5) There is some (more or less indirect) evidence which supports the X-ray, rather than the oil-drop, result. It is perhaps scarcely strong enough to justify, and certainly not strong enough to compel, an immediate decision in favour of the higher value of e.

ADDENDUM

In a very recent paper[54] (September, 1935), amplifying an earlier note[31], Bearden gives the details of his latest work, in which the possibilities of the ruled grating in X-ray spectrometry have been very fully exploited. Bearden compares his own results with those of Bäcklin[29] and Södermann[30], reducing all values to a common crystal wave-length scale. He finds, as the weighted mean of about 850 measurements (which include one set of refraction experiments[50]), that the excess of the ruled grating and refraction values over the crystal values of the wave-lengths used is $(0{\cdot}248 \pm 0{\cdot}0016)$ per cent. The crystal wave-lengths are given in terms of Bearden's own value[11] of the calcite constant, which depends on the assumption that e is $4{\cdot}768 \times 10^{-10}$ e.s.u. Bearden's final value for e is therefore $(1{\cdot}00248)^3 \times 4{\cdot}768 \times 10^{-10}$ or $(4{\cdot}8036 \pm 0{\cdot}0005) \times 10^{-10}$ e.s.u.

REFERENCES

(1) Birge. *Phys. Rev. Suppt.* (now *Rev. Mod. Phys.*), **1**, 1 (1929).
(2) Millikan. *Phys. Rev.* **35**, 1231 (1930).
(3) Birge. *Nature*, Lond., **133**, 648 (1934).
Bond, *Nature*, Lond., **135**, 825 (1935).
(4) Townsend. *Proc. Camb. phil. Soc.* **9**, 244 (1897).
(5) —— *Phil. Mag.* **45**, 125 (1898).
(6) Thomson, J. J. *Phil. Mag.* **46**, 528 (1898); **48**, 547 (1899).
(7) Moseley. *Phil. Mag.* **26**, 1024 (1913).
(8) Millikan. *Phys. Rev.* **32**, 349 (1911).
(9) Compton, Beets and De Foe. *Phys. Rev.* **25**, 625 (1925).
(10) Siegbahn. *Phil. Mag.* **37**, 601 (1919); *Ark. Mat. Astr. Fys.* **21** A, No. 21 (1929).
(11) Bearden. *Phys. Rev.* **38**, 2089 (1931).
(12) Thomson, J. J. *Phil. Mag.* **5**, 346 (1903).
(13) Wilson, H. A. *Phil. Mag.* **5**, 429 (1903).
(14) Millikan and Begeman. *Phys. Rev.* **26**, 198 (1908).
(15) Millikan. *Phil. Mag.* **19**, 209 (1910).
(16) —— *The Electron.* [New Edition: *Electrons* (+ *and* −), *Protons, Photons, Neutrons and Cosmic Rays* (Cambridge University Press, 1935).]
(17) —— *Phys. Rev.* **2**, 109 (1913).
(18) —— *Phil. Mag.* **34**, 1 (1917).
(19) Birge. *Phys. Rev.* **40**, 228 (1932).
(20) —— *Phys. Rev.* **42**, 736 (1932).
(21) Compton and Doan. *Proc. nat. Acad. Sci.*, Wash., **11**, 598 (1925).

(22) WADLUND. *Phys. Rev.* **32**, 841 (1928).
(23) BÄCKLIN. Upsala Dissertation (1928): *Nature*, Lond., **123**, 409 (1929).
(24) COMPTON, A. H. *J. Franklin Inst.* **208**, 605 (1929).
(25) CORK. *Phys. Rev.* **35**, 128, 1456 (1930).
(26) SHIBA. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **19**, 97 (1932); **21**, 128 (1933).
(27) WITMER and CORK. *Phys. Rev.* **42**, 743 (1932).
(28) BEARDEN. *Phys. Rev.* **37**, 1210 (1931).
(29) BÄCKLIN. *Nature*, Lond., **135**, 32 (1935); *Z. Phys.* **93**, 450 (1935).
(30) SÖDERMANN. *Nature*, Lond., **135**, 67 (1935).
(31) BEARDEN. *Phys. Rev.* **47**, 883 (1935).
(32) SIEGBAHN. *Spektroskopie der Röntgenstrahlen* (1931). Berlin: Springer.
(33) COMPTON, A. H. and ALLISON. *X-Rays in Theory and Experiment* (see especially Chapter 9) (1935). Macmillan.
(34) DARWIN. *Phil. Mag.* **27**, 315, 675 (1914).
(35) PORTER, A. W. *Phil. Mag.* **5**, 1067 (1928).
(36) TU. *Phys. Rev.* **40**, 662 (1932).
(37) WILSON, H. A. *Phys. Rev.* **34**, 1493 (1929).
(38) HARRINGTON. *Phys. Rev.* **8**, 738 (1916).
(39) VAN DYKE. *Phys. Rev.* **21**, 250 (1923).
(40) RUTHERFORD and GEIGER. *Proc. roy. Soc.* **81**, A, 162 (1908); *Phys. Z.* **10**, 42 (1909).
(41) REGENER. *Berl. Ber.* **2**, 948 (1909).
(42) RUTHERFORD CHADWICK, and ELLIS. *Radiations from Radioactive Substances* (1930). Cambridge University Press.
(43) BRADDICK and CAVE. *Proc. roy. Soc.* **121**, A, 367 (1928).
(44) WARD, WYNN-WILLIAMS and CAVE. *Proc. roy. Soc.* **125**, A, 713 (1929).
(45) SCHOPPER. *Z. Phys.* **93**, 1 (1934).
(46) BIRGE and MCMILLAN. *Phys. Rev.* **47**, 320 (1935).
(47) SCHOPPER. *Z. Phys.* **94**, 649 (1935).
(48) BEARDEN. *Phys. Rev.* **39**, 1 (1932).
(49) BEARDEN and SHAW. *Phys. Rev.* **48**, 759 (1934).
(50) BEARDEN. *Phys. Rev.* **47**, 645, 811 (1935).
(51) KRETSCHMAR. *Phys. Rev.* **43**, 417 (1933).
ROBINSON, ANDREWS and IRONS. *Proc. roy. Soc.* **143**, A, 48 (1933).
ROBINSON. *Phil. Mag.* **18**, 1086 (1934).
See also other papers quoted in ROBINSON and CLEWS. *Proc. roy. Soc.* **149**, A, 587 (1935).
(52) RUARK. *Phys. Rev.* **47**, 316 (1935).
(53) SHANE and SPEDDING. *Phys. Rev.* **47**, 33 (1935).
(54) BEARDEN. *Phys. Rev.* **48**, 385 (1935).

ELECTRON TUBES

Edited by Prof. G. I. FINCH.

PART I. GENERAL INTRODUCTION

By Prof. G. I. FINCH.

THE general term "electron tubes" is here intended to cover those instruments in which fine beams of electrons moving at uniform velocity are deflected for various purposes either by electrostatic or magnetic fields, as in the cathode-ray oscillograph and electron microscope, or are scattered by atomic aggregations as in the electron diffraction camera. In the case of the oscillograph we ostensibly rely on the charge-mass properties of the electrons for the interpretation of the phenomena, whilst in electron diffraction, owing to the nearly nuclear localization of the deflecting fields, it is the wave aspect of the moving electrons which is most in evidence.

Owing to the high value of the electronic charge/mass ratio, even at the readily attainable extremely high electron speeds now so frequently employed, the electron beam is easily deflected by transverse electrostatic and magnetic fields of moderate intensity and can follow rapid fluctuations therein with remarkable faithfulness. Furthermore, coaxial symmetrical fields affect the electron paths in a manner somewhat analogous to that of an optical system in the case of light, and since the wave-lengths associated with electrons of quite moderate speed are some 6×10^4 times shorter than those of visible light, the possible resolving power of such electron-optical systems is correspondingly high. The approximately 10^7 times more intense nuclear interaction of electrons as compared with the extra-nuclear electronic scattering of X rays, and the consequent poor penetration by even fast electrons results in the scattering effects being virtually confined to the surface layer of the material; accordingly, the electron diffraction camera has become the instrument *par excellence* for the study of the crystalline structure of thin films and surfaces.

The cathode-ray oscillograph has developed rapidly in the past ten years, chiefly owing to the demands made upon it by the radio, television, high-tension electrical and switch-gear industries, and has now reached a stage where such high writing speeds are readily attainable as involve correction of the oscillograms for the relativistic mass changes of the deflected electrons.

When G. P. Thomson, eight years ago, first put de Broglie's wave theory on an experimental basis, he was quick to foresee the potentialities of his discovery as a tool for the study of surface structure problems, and such rapid progress has since been made that electron diffraction is now widely used as a method of structure analysis.

Busch's classical work in 1926 on the focussing of electron beams laid the foundation of electron optics and soon led to the construction of the electron microscope, clearly destined before long to fill the gap between the optical microscope and the electron diffraction camera.

Part 2. THE HIGH-SPEED CATHODE-RAY OSCILLOGRAPH

By J. L. Miller, D.Eng., Ph.D., M.I.E.E., Mem.A.I.E.E. and J. E. L. Robinson, M.Sc.

§ 1. INTRODUCTION

From the first production of Geissler discharges and through the early experiments of Sir J. J. Thomson to determine the values of e, e/m and v for cathode rays and the early attempts in the same year of Braun to apply cathode rays to commercial measurement, the history of the cathode ray oscillograph is admirably traced along the many stages of its development up to about 1925 by Prof. MacGregor-Morris and R. Mines[1] in a paper published in the *Journal of the Institution of Electrical Engineers* in that year.

Just before that time the cathode ray oscillograph having the widest application —at any rate in this country—was the familiar small sealed-off instrument* of the Western Electric Company, but this and, of course, the more recent instruments of the same type do not come within the scope of the present article since they are not suitable for the recording of short-time single transients.

The need for measurements of this type in electrical engineering caused Dufour[2] to have developed a completely new type of oscillograph as early as 1922. Differing from its predecessors in many ways, the two most important innovations were the use of a high accelerating potential of 60,000 V. on a cold cathode and the placing of the photographic plates and films inside the vacuum, and due both to the greatly increased energy of the electron beam and to the fact that it acted directly on the photographic plate, the recording sensitivity was increased enormously. This oscillograph, with the exception of the discharge tube, was made of metal and this together with the fact that the photographic plates gave off vapour and also required replacement necessitated continuous evacuation which was effected by means of a mercury diffusion pump, while air was permitted to leak in through an adjustable needle valve to provide the internal gas pressure requisite for the cold discharge.

In 1923–5 Wood[3] also developed a continuously evacuated instrument which employed a hot cathode and an accelerating potential of 3000 V., and consequently in 1925 there were in existence two types of oscillograph which might be termed forerunners of the modern high-voltage high-speed oscillograph. No technique, however, was available for the measurement of transient voltages of durations of a few microseconds only, and no doubt if it had not been for a very practical impetus

* Operating with 300 V. cathode potential and providing a trace on the internal fluorescent screen most suitable for visual examination, these low voltage tubes and, of course, the more recent ones of the same type are very suitable for recording repeatable or continuous oscillations where a visually steady trace is obtained on the screen by arranging the time sweep motion so that the spot is rapidly retraced synchronously with the phenomena under observation. Their focussing action is, however, dependent on ionization of the internal gas and sluggishness of this mechanism leads to a loss of definition and unfaithful response above a frequency of the order of 200,000–500,000 cycles per second, so that they are not strictly high-speed instruments. In any case, it is probable that they will be superseded for many purposes by the sealed off modern electrostatically focussed high vacuum oscillographs whose frequency response is not similarly limited and which do not suffer from the non-linear deflection calibration also a characteristic of the gas-filled type.

a considerable number of years would have elapsed before an accurate and reliable instrument together with a well established technique had been evolved.

From about 1924 onwards, however, overhead lines operating at pressures up to 220 kV. were being erected and with the increasing numbers of these exposed systems the interruptions of service and the damage to connected equipment resulting from lightning causes rapidly became of great importance. Measures for the overcoming and prevention of the troubles were hampered by lack of understanding of the mechanism of their occurrence and the realization of this demonstrated the great need for apparatus for measurement of the lightning transient over-voltages.

And thus it was that the Dufour oscillograph had reached a useful stage of development just coincidently with the arising of an insistent demand for such an instrument and was ready at hand for adaptation by American engineers[4, 5] to the recording of the lightning phenomena on the overhead lines.

At about the same time the evolution of oscillograph instruments and technique was proceeding on the Continent due to the efforts of many engineers including Gábor[6], Knoll[7], Krug[8], Norinder[9], Rogowski[10] and Berger[11], but here contrasted with the conditions in America the development was taking place in the more tranquil academic atmosphere and the trend centred more in the perfecting and elaborating of the instruments rather than in a hurried application to very high voltage work in the field.

In this country the need for high-speed transient measurement had been early felt in the Metropolitan-Vickers[12] and Ferranti[13] Laboratories, and these concerns were the first here to build high-voltage oscillographs and since these have been intended primarily, at any rate initially, for research in the laboratory their development has been characterized chiefly by the requirement of great accuracy* in high voltage measurements. A little later the high-speed instrument of Finch[15], of which one notable installation by the Cambridge Instrument Company is that at the National Physical Laboratory, has made its appearance, and it is of interest to note that Finch's instrument has been developed jointly for the requirements of oscillography and electron diffraction measurements.

As a consequence to-day the high-speed oscillograph owes its advanced state of development and importance in no small measure to the combined needs and efforts of the transmission engineer and of the manufacturer of transmission equipment and its main field of application still is that of high-voltage short-time impulse phenomena, where the technique has so far advanced that photographic recording in Cartesian coordinates can be made with good accuracy of transients whose duration is of the order of microseconds even up to the highest pressures of several million volts.

In addition, however, to the main high-speed type, mention must be made of one other hybrid apparatus which is very similar to the original Dufour type and of which several examples have appeared since 1933. This instrument which employs

* Both Allibone and the authors have recently referred[14] to measurements of high transient voltages with an error not exceeding ±1·5 per cent.

the normal cold cathode discharge tube, focussing coil, trap system and continuous evacuation of the high-speed oscillograph in conjunction with a mechanical time delineation motion in the form of a rotating drum inside the vacuum has been developed for investigating the performance of electrical apparatus under power frequency short-circuit conditions where the simultaneous existence of the slow power frequency and the fast transient frequencies entails the use of records of great linear length.

§ 2. GENERAL CONSTRUCTION

In nearly all high-speed oscillographs the use of internal photography is general and on this account at least continuous evacuation, which renders readily possible the usual metal construction, is employed. The discharge tube and electrode bushings are thus the only insulators required and these are variously of glass or porcelain cemented in suitable fashion on to the metal construction. A vertical oscillograph design is the most popular although several workers employ a horizontal layout, in which connexion a minor point is the greater compensation necessary to counteract the effect of the stronger vertical component of the earth's magnetic field.

Cold cathode discharge tubes operate at a pressure of the order of 10^{-3} mm. of mercury, while the requirements of no scattering of the subsequent beam and freedom from conduction of the deflection electrodes tend to require a lower gas pressure* in the remainder of the apparatus. In continuously evacuated apparatus maintenance of the correct pressure in the discharge tube is observed by pumping here against a controllable air leak. In many designs the discharge tube is evacuated through a small channel in the anode leading into the lower part of the apparatus and across this the required pressure difference is maintained by the single pumping equipment attached to the main housing, generally in the region of the photographic chamber. It is the practice in others to employ an additional pumping unit exhausting the discharge tube directly and here the leakage to the main pump from the discharge tube through the small anode aperture is slight. In this case, of course, smaller bore pumps can be used, but the duplication of equipment appears to have little advantage in general although it is probably of assistance in special designs in obtaining the lower pressure necessary for the operation of deflecting electrodes at high potentials to ground. The deciding factor in all cases is chiefly that of the pumping speed required and the capacity of pumping equipments in use on oscillographs ranges from about 5–20 litres per second.

The type of vacuum equipment employed is always a matter of choice. Mercury diffusion pumps were early used and are still widely favoured, the use of vapour traps being unnecessary (with cold-cathode instruments) for the degree of vacuum required, at any rate with the normal deflection plates designed for a few kilovolts pressure only. Oil diffusion pumps employing the oils isolated by C. R. Burch[17] are notably used by Metropolitan-Vickers[12], while in some American equipments

* In the sealed-off cold cathode instrument of the Rogowski School the same gas pressure, of course, obtains throughout the apparatus[16].

rotary molecular pumps appear to be favoured. Diffusion pumps are either of single or a greater number of stages according to design and are backed by the usual rotary rough vacuum pumps. Two-way taps in the pumping line can be employed to allow preliminary evacuation of the oscillograph by the mechanical pumps only and thus permit the diffusion pumps to be kept running continuously so that after reloading the camera a few minutes suffices to re-exhaust the apparatus.

The leak for controlling the discharge tube pressure is usually of the adjustable needle valve type (leaking either from atmosphere or the fore-vacuum supply) although an ingenious exception is that by Whipple[18] where leakage is through a fine porous porcelain whose cross section is varied by a mercury seal. One sealed-off version of the cold cathode tube exists due to the Rogowski School[16] and here the control of the gas pressure is obtained by heating a strip of palladium in the vacuum to drive out occluded gas and make up the deficiency as required.

Rigid jointing of the vacuum parts is variously effected by welding, brazing, soldering and cements or waxes according to the permanent or temporary nature of the construction. Where operating movements require to be carried through the vacuum wall these are effected in the case of rotary motions, such as for camera operation, by conical or more popularly (at any rate in this country) by flat ground joints lubricated and sealed by tap greases. For adjustments such as of deflection electrode separation tombac flexible metal connexions are sometimes used. For the removable camera cover plate a greased rubber seal is usually employed, while for demountable pieces, such as those required for cathode renewal, wax or plane greased joints are in common use.

In the internal rotating drum oscillographs a high speed rotary motion of course requires to be carried into the vacuum. In the instruments of Van Siekle[19] and Whelpton[20] this is effected by the use of an induction motor drive with the rotor inside and the stator external to the vacuum, while in the case of the similar instrument of the Cambridge Instrument Company[20] the drum is driven by an external motor, the shaft entering from atmosphere through a system of glands and a chamber kept at a pressure intermediate between that of the photographic chamber and atmosphere.

§ 3. DISCHARGE TUBES

With a few exceptions discharge tubes are of the cold cathode type employing aluminium cathodes and operating at various voltages from 40 to 60 kV. In the present state of knowledge of gaseous discharge at the lower pressures, their design is largely empirical and as a consequence most laboratories have developed their own tubes along individual lines. The main requisite of a discharge tube for a high-speed oscillograph is the ability to produce continuously a steady and concentrated beam from a small emitting area. At the high necessary voltage this beam cannot be maintained by the use of a simple electrode system such as there is in an ordinary Geissler tube and, therefore, satisfactory designs require the use of insulating or metal shields and suitably shaped electrodes.

The original oscillographs employed glass discharge tubes and such are still used by Finch and many American workers and, in general, by the Aachen Institute*[21, 22]; in fact, some of the fastest examples of high-speed tracing were obtained by Rogowski[26] by their means. In their simpler forms, however, as mentioned by various people, they are liable to flickering and heating troubles—which do not matter so much in the case of the American workers who only energize their cathode during photographing—and the eradication of these has led to the

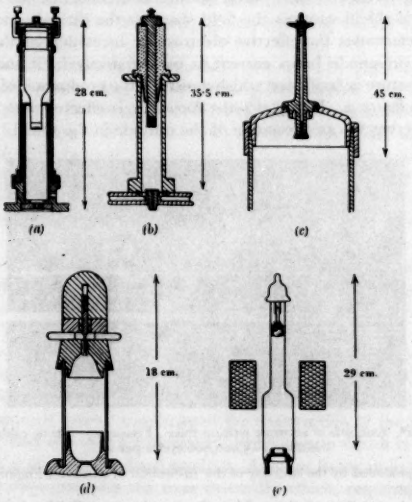


Figure 1*. Discharge tubes employed by various workers showing approximate dimensions.

* (*b*), (*d*), (*e*) reproduced by courtesy of the Institute of Electrical Engineers. (*a*) and (*c*) reproduced by courtesy of the Elektrotechnische Zeitschrift.

evolution of designs employing a minimum amount of insulation in the construction. Such tubes can be made very stable in operation and, with the provision of water cooling of the anode, will run steadily for long periods.

One of the earliest metal discharge tubes was that due to Knoll[27] which is shown in figure 1*a*, but Burch and Whelpton[12] referred to difficulties in operating this tube, and they accordingly developed a simpler design. This tube, shown in figure 1*b*, is still employed in the Metropolitan-Vickers' Laboratories, and its

* J. M. Dodds, of the Aachen Institute, which School has predominantly used glass tubes, has employed a tube of the Burch and Whelpton type for recording phenomena at very high writing speeds using external photography[23]. Dicks has discussed the pros and cons of glass and metal tubes[24] and one of the authors (J. L. M.) has also commented on this question[25].

characteristics are such that Whelpton has been able to operate with complete stability three such units in parallel at 15 kV. for his rotating drum oscillograph[20].

Details of a rather different form of tube, in which pre-anode electrostatic focussing of the ray was obtained by surrounding the cathode with a concentric metal shield, have been published by Binder[28] and this tube, without its shield, is shown in figure 1*c*.

The authors[29] operate the tube shown in figure 1*d*. While glass is employed as the necessary insulation, it is purely as such and is outside the main discharge field. The metal shield controls the field shape at the cathode and consequently its potential determines the effective electrostatic focussing, so that the ratio of post-anode to pre-anode beam current is comparatively high and the use of a small anode aperture is facilitated which assists in the production of the thin dense trace shown in figure 2. In practice the focussing is effected with floating shield potential by varying the axial position of the cathode in the shield.

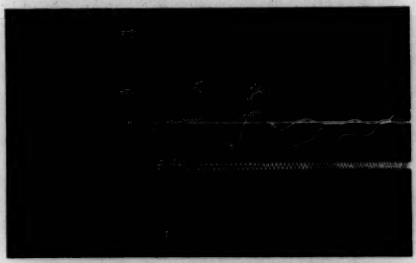


Figure 2*. Example of accurate writing trace. Frequency of time calibration oscillation = 5,000,000 cycles per sec.

* Reproduced by the courtesy of the Institution of Electrical Engineers.

The ready adaptability of the cold cathode tube is well illustrated by the suggestion of Knoll[30], who proposed the use of a wide sectioned beam and a four-aperture anode with four separate deflection systems, so that four separate phenomena could be photographed simultaneously. Recently Boekels and Dicks[31] have produced such a tube, so that two cathode-ray counterparts of the multi-element Duddell oscillograph, this high-speed one and in addition the lower speed arrangement of Whelpton, are now in existence.

Nearly all cold-cathode oscillographs are continuously evacuated by high-speed pumps and hence the renewing of a worn cathode is simply carried out. However, a recent 40 kV. cold cathode instrument, illustrated in figure 1*e*, due to the Rogowski School and described in England by Parker Smith and others[16], is sealed off, and in this case the effect of wear of the cathode is ingeniously countered by designing it in the form of an aluminium sphere which can be shaken so as to present a new emitting surface as required.

An interesting phenomenon observed in the operation of metal discharge tubes is the production of retrograde rays, the intensity of which appears to vary with the design of the discharge tube. Not being deflected by electric or magnetic fields, they impinge upon the photographic plate and cause some fog, and this may become important when using continuously energized cathodes for field work. They can be avoided by filtering the ray through thin aluminium foil in the oscillograph, by using the Norinder target, or by tilting the discharge tube and returning the ray to the oscillograph axis by an external field in the manner of Whelpton[20]. Finch, on the other hand, also tilts the trap plates, returning the ray to the axis by means of the focussing coil[20].

The limit of faithful recording of ultra-fast phenomena by high-voltage instruments using inside photography (apart from potentiometer and circuit considerations) occurs when the time taken for an electron to pass a deflection plate* is comparable with the equivalent periodic time of the recorded phenomenon and is not due to any lack of intensity of the trace, and in fact in Germany several workers have produced oscillograms having recording speeds of the order of one fifth of the velocity of light. These very high-speed recordings appear to have been made with cold-cathode tubes and generally with the provision of pre-anode electromagnetic focussing. Probably only Rogowski and Ackerman[33] use this last device for normal high-voltage impulse work since the very maximum writing speed generally required is only of the order of one-hundredth of the velocity of light and with internal photography and the usual discharge tubes all ordinary recording can be made without its assistance.

The fact that cold cathode tubes have been so extensively employed in preference to hot cathode ones may appear surprising, particularly when it is realized that the former have long passed from favour in the X-ray field. Contributory reasons are the inherent greater simplicity of the construction of the gas discharge tube and the fact that the very small cathode emitting area which is required for oscillograph work is almost its natural product. Another factor in its favour is the usual practice of photographing the trace internally which, requiring continuous pumping plant, renders simple the maintenance of the required critical gas pressure. The hot-cathode continuously evacuated tubes so far constructed do not appear to have any advantage over the cold cathode tubes, and the latter, from the points of view of simplicity and fineness of trace, appear superior. One notable hot-cathode tube, that due to George[34], is in use in America and lends itself very well to the frequent American practice of cathode energization only during the period of photographing by virtue of the fact that the ray may be switched on by application of voltage to the control electrode of the tube *via* a valve relay. Very early Rogowski and his associates also operated high-voltage hot cathode tubes[35] and recent improvements have been suggested by Graupner[36] in a design which incidentally included a canted metal discharge tube. Another high-voltage hot cathode instrument has recently been described by Gondet and Beaudouin[37].

* Klemperer and Wolff have discussed the effect of the dimensions of the deflecting system on the percentage distortion at high writing speeds[1, 32].

Oscillographs with a high accelerating voltage have the disadvantage for some purposes of a comparatively low deflection sensitivity. It is interesting to note, therefore, that several instruments employing low accelerating voltages have appeared, and one such apparatus with an accelerating voltage of 6 kV. (using a cold cathode and pre-anode electromagnetic focussing) enables frequencies as high as 1 megacycle to be photographed with an external camera in a single traverse[38].

Another interesting innovation enables the desirable photographic sensitivity of the higher voltage oscillographs to be retained in part while yet maintaining the deflection sensitivity of the low voltage instrument. Here pressures of the order of 1 kV. are employed between cathode and anode, and the main accelerating potential of several kilovolts is applied after the co-ordinate deflection. Such an instrument, using a hot cathode and internal photography, was early described by Sommerfeld[39] in 1928 and with only 150 V. on the cathode but a total accelerating voltage of 5000 V. a voltage sensitivity of 0·7 cm./V. was obtained compared with the figures of 0·025 cm./V. obtained by Wood using 3000 V. on the cathode and 0·001 cm./V. approximately for the normal high-voltage oscillograph. Graupner's oscillograph is virtually an improvement on another one due to Rogowski, Sommerfeld and Wolman[40] where light-blackening contact photography (the photographic plate being in a chamber at fore-vacuum pressure, thus enabling unsupported glass to be employed) was used, having a greater voltage deflection sensitivity. Harris[41] has also produced an instrument in which two stages of electron acceleration are used.

Whether or no separate pre-anode focussing is employed, all oscillographs require a further focussing device between the anode and the recorder and for high-voltage instruments this has generally taken the form of an iron shrouded magnetic concentrating coil having about 1500 A. turns in the case of 50 kV. instruments. An analysis of the action of this apparatus by Bush[42] produced the conclusion that it was desirably mounted equidistant from anode and recorder, but from observation of existing instruments it would appear that all workers do not rigidly adhere to this requirement, and it has been the authors' experience that considerable latitude is permissible.

More recently, the electrostatic method of focussing[43,76] using coaxial apertures with controlled potentials has been developed and applied to some high-voltage tubes, but at any rate so far it is not widely employed.

§ 4. PHOTOGRAPHY

Internal photography is most widely and popularly employed probably both because of its simplicity and the high definition which can be obtained at all writing speeds. The use of both plates and films is general, the former having the advantage that if necessary only a part of the batch need be removed at a time for development with consequent avoidance of waste, while the latter have a slight disadvantage of greater time required for evacuation. The action of the electron beam direct on the photographic emulsion is the most efficient method of recording and at any

rate for normal purposes slow emulsions are just as sensitive to electron action as the faster ones. The highest recorded writing speeds have been made with internal photography, one as high as one-fifth the velocity of light having been attained by Rogowski[44].

External photography, which of course has to be employed with sealed off instruments, has the obvious advantage of not disturbing the vacuum at recharging of the camera. Its development for high voltages has been mostly confined to the efforts of German workers and three distinct methods, the external electron blackening, the light blackening contact and the lens and camera methods are in use.

In the first of these the Lenard window has been variously made of aluminium* and cellon foil, these two low density materials being about the best in relation to their mechanical strengths for transmitting high speed electrons. With the thinnest foils so far successfully employed, however, reasonable photographic efficiency is only obtained by the use of very high cathode voltages and the foils being slightly porous have so far precluded a sealed-off construction on this principle which is, of course, one of the main advantages to be sought from external photography. No attempt appears to have been made with this method to reach the limit of its recording speed but one of one-sixtieth that of light has been quoted by Knoll[15, 46] and a speed not far short of that of internal photography under similar conditions with high accelerating potentials appears to be practically possible. As regards definition this method gives about the best of the external ones and may not be much inferior to that of internal photography.

The light blackening contact method, which is incidentally the only external method which the authors[13, 29] themselves have used, is dependent for its efficiency on the use of a thin window (0·2 mm.) and on the preparation of the fluorescent screen and its material, there being an optimum thickness of coating for each cathode voltage. Several fluorescent materials are employed but zinc sulphide (of certain brands of manufacture) appears to be one of the most suitable for high voltage rays and with such a screen the photographic efficiency is stated to be as high as two-fifths[47] of that of internal recording under the same conditions, although the limit of recording speed reached is not stated.† A thickish trace of comparatively poor definition is, however, characteristic of the published records made so far by this method and no sealed off construction of this type has yet appeared.

The third and readiest external method with lens and camera lends itself to observation of the free surface of the fluorescent screen and the design of the latter is, therefore, more simple. A slight point in connexion with this method is the fact that in any case in all oscillographs such a fluorescent screen must be provided for visual observation. At normal writing speeds records with quite good definition are obtained by this method, and Dodds[23] employing 90 kV., a continuously evacuated oscillograph and a high beam current value with pre-anode focussing has recorded a writing speed of one-tenth that of light, being half as rapid as the fastest

* For a discussion on window thickness see ref. [77].

† Rogowski has also discussed the present position of the several recording methods, see ref. [48].

internal recording, although a photographic reduction of about fourfold from the screen to recording plate entails the speed of the actual recorded trace being again a quarter this value. So far, the published oscillograms taken by this means have been typified by a considerable variation of thickness of trace from point to point on the same record for even relatively small variations of writing speed, so that under adverse conditions its definition has been inferior to that of the internal method. The use of high accelerating potentials is not, however, essential to this method, normal writing speeds having been recorded with a cold cathode instrument[38] with as low as 6 kV. potential. The light emission of the screen is most efficient on the beam side and for this reason various devices have to be adopted for obtaining an undistorted picture in the external camera, such as tilting the screen towards the camera or off-setting the camera lens from the photographic film.

Contrasted with internal photography the several external methods have some disadvantage in that they are less foolproof and this is the more so in the cases of the foil and glass window contact methods which are readily punctured by a stationary beam, while in the case of the last method the screen is liable to burning and loss of sensitivity under the same condition. With continuously evacuated instruments replacement of course is simple and in the sealed off lens and camera system where the possible damage is not so severe this cannot be considered an important disadvantage.

Although the external methods may not permit the obtaining of quite the same limiting recording speeds as do the internal methods, for nearly all practical purposes their sensitivity can be amply high and it is really chiefly on the score of their varying degrees of inferiority of definition that they compare unfavourably with internal photography. It must, however, be borne in mind that any system of external photography is simpler in conjunction with continuously evacuated equipment by reasons of robustness and ease of replacement. With a sealed off construction the apparatus cannot normally be forced to the same limits of performance. In the present state of development there is little doubt that for normal purposes the choice of a method of recording lies only between the use of internal photography and external photography with lens and camera, and it is probable that considerable development of the latter type of apparatus will take place.

§ 5. BEAM RELEASE AND WITHDRAWING SYSTEMS

The most important cathode beam switching devices which are employed to prevent the ray lingering on the photographic plate before and after recording may be classified in two groups corresponding to whether the beam is switched at the cathode or immediately after the anode.

The first system, introduced by Dufour[2], applies exciting voltage to the discharge tube for a predetermined duration sufficient for recording. Adapted by a spark gap switching device (see figures 6 and 8) to automatic control by the arriving transient this method is employed by many American workers[49, 50, 51, 52]. While it is very effective for long period field work the authors have no experience

of it, and it is not universally employed in America and not at all in Germany and this country.

The second system which is almost universally used utilizes an electromagnetic or electrostatic control in a separate trapping chamber acting directly on the electron beam after it emerges from the anode.

An electrostatic method early developed by Norinder[9, 33] in its simplest form has two pairs of deflecting electrodes successively arranged along the oscillograph axis and cross-connected so that any potential applied to them, having the deflection of the first pair exactly neutralized by that of the second, produces no resultant motion of the recording beam. Between the pairs, however, the ray moves laterally to one side and a small target placed there on the axis serves to block its path except when voltage is applied to the electrodes. In this way the device is suited for uncontrolled transient work, the application of any unidirectional impulse to the plates releasing the beam for just so long as its duration. This method has been used notably by the American Westinghouse Company[51] (in an elaborated form) and by Norinder in Norway. It is very effective in operation, while

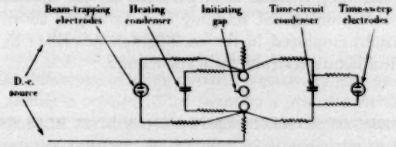


Figure 3. Example of oscillograph auxiliary basic circuit.

its chief disadvantages are that its operating time is controlled by the impulse wave shape and that it is dependent on the latter being unidirectional.

Originally this trap system was employed with an oscillatory time motion[50, 53], but nowadays, owing to the trouble of replotting the results, the relay is used simply as a beam trap and the surge is made to initiate a normal single sweep time base.

The other electrostatic method, developed originally in Germany, fundamentally employs only one pair of deflection electrodes and as contrasted with Norinder's method releases the beam by permitting it to pass through a central aperture when the applied voltage drops to zero. Its operation, therefore, is controlled by a local auxiliary circuit, the deflection plates being connected across the spark gap (as in figure 3) which controls the main time sweep circuit and thus inherently releases the beam in about the time of growth of the spark when the gap discharges.

A modification of this method (Gábor) provides an additional pair of electrodes cross-connected from the first and following the blocking diaphragm; it is thus virtually a complete inversion of the Norinder relay, the aperture replacing the target. Giving as it does no resultant deflection for any stray oscillation in the operating circuit, this system perhaps requires rather less care than the previous in the layout of the connexions of the associated circuits, but there is little to choose

between these two methods. The former is favoured by, amongst others, Krug, Binder, Finch and the authors, while the latter, amongst its devotees, numbers Whelpton.

With these zero voltage release systems withdrawal of the ray at the completion of time-sweeping is automatic, the electrode voltage commencing to rise immediately the spark gap extinguishes. The ray can thus be withdrawn before recharging of the time sweeping condenser begins to retrace the beam across the plate and withdrawal by this means may take, at the most, of the order of 100 microseconds, which is quite sufficiently brief when, as is usual, the time-sweep motion carries the beam off the recorder.

Authorities[54] in Germany have suggested other means of effecting withdrawal; Rogowski a Faraday cage which causes the beam to retrap itself, and Boekels an electromagnetic withdrawing circuit whose action *via* a three electrode valve is delayed until the circuit voltage has fallen to a predetermined value. These methods, however, do not appear to have effected any useful purpose and they are certainly not in general use. Other trapping systems for use with controlled transients have been suggested, and they include such methods as the mechanical shutter of Rogowski and the electromagnetic trapping coils (arranged above the anode and operated by a switch) employed in the oscillograph described by Parker-Smith, but their mechanical nature restricts their application.

§6. METHODS OF TIME DELINEATION AND BIAS CONTROL

For time delineation Dufour[2] in his early work used methods ranging from the rotating drum to those superimposing several high frequency oscillatory motions on the beam. Although the simple oscillatory time motion was favoured by some up to recent times (and perhaps in some instances still is) in conjunction with the Norinder relay, nowadays for single transients persisting for a fraction of a microsecond to a hundred or so microseconds, following Gábor, single sweep time motions are employed.

The mechanical method, unsuited to normal high-speed transient measurement as mentioned previously, has recently come into prominence for medium frequency work. This work, dealing with such phenomena as the recovery voltages in circuit breaker testing, demands records covering several complete 50-cycle waves on a time scale such that frequencies of the order of 100,000 cycles per second can be delineated. It is quite evident, first, that single sweep time bases would be incapable of dealing with these requirements, and secondly, that the frequencies involved are well beyond the range of the Duddell oscillograph. Therefore a solution is to employ a rotating drum attachment to a normal high-speed oscillograph, and since 1933 three such instruments have been produced.

The requirement of a record of very great linear length has been satisfied in a novel manner by Whipple[20]. The drum is allowed to rotate five times during recording and the five traces so made are separated by traversing the beam transversely across the film after each revolution so that with a maximum drum speed

of 3000 r.p.m. a complete 50-cycle period occupies 10 metres total length of trace. In order to do this a pair of zero-shifting deflection plates, fed with potentials from a commutator geared to the camera drive, is incorporated in the deflection chamber. To increase still further the length of the record, Harle and Cassie[20] suggested the application of a relatively slowly varying potential to the zero-shifting electrodes, and in this way records as long as 50 metres were obtained.*

In another instrument by Whelpton[20] such great linear length of record was not required. At a drum speed of 3000 r.p.m. one complete 50-cycle wave occupied the 1 metre total length of film. This instrument, however, is notable in that by the provision of three discharge tubes and three trap and deflection systems mounted on the same photographic chamber three records can be made side by side on the same film, thus making it suitable, when necessary, for three-phase working.

In the third instrument of that Van Siekle and Berkey[19] of which details were the first to be published, only single phase operation was aimed at and a maximum drum speed of 7200 r.p.m. sufficed to delineate half a 60-cycle period across the 0·46-metre length of film, while greater linear lengths of trace were here obtained merely by superposition during several successive revolutions of the drum made asynchronously with the power supply.

Contrasted with the continuously operating type of time motion the single-sweep one has the difference of being initiated electrically only on the advent of the transient. This motion is applied to the beam by a pair of deflection electrodes fed with the exponentially varying voltage of a condenser-resistance circuit whose charge or discharge is initiated by tripping of a spark gap or by a valve relay. Most generally in Germany and this country the action at discharge is employed, while in America† the charging action is frequently favoured, and usually the oscillograph deflection plates are connected across the condenser.‡ The type of action employed is, of course, bound up with synchronizing methods and bias considerations.

Some workers care to substitute a linear variation with time for the exponential one by employing a saturated diode as discharge resistance. Some complications in its use at very high speeds and the inconsistency of its calibration contribute, however, to a general preference for the first method with the active use of perhaps only half the condenser voltage fall to give a sweeping motion usefully approximating to linear. In any case non-linearity causes no inconvenience since it is current practice to superimpose on each oscillogram a small amplitude oscillation of convenient periodic time such as 1 microsecond in order to provide an accurate time calibration.

It is impossible to deal with the large number of circuits in use, but a basic

* This rotating drum camera may be removed and substituted by a camera suitable for plates or films and the zero-shifting deflection plates can be transferred to a position 90° removed from their previous one, so that the instrument is also suitable for high-speed single-transient work.

† See for instance [52, 55, 56].

‡ Krug [57, 58], amongst others, has discussed charging and discharging time circuits and their connexions, including one, due to him, where the ray is made to arrive on the recorder with a preliminary rapid leap.

connexion employing the discharging action of a condenser and a spark gap relay, which serves to illustrate in its essentials the practice followed by most European laboratories for high-voltage work, is shown in the right-hand side of figure 3. Several workers[12, 20] have referred to the advantage of maintaining the potential of the time-sweep electrodes always equally divided about earth and to this end the discharge resistance is shown equally divided (as has been standard practice for some years in Germany) about the spark gap.

The beam-trapping electrodes are shown connected across the outer spheres, although they are often suitably connected across the middle sphere to ground, and thus this circuit is also automatically initiated on sparkover of the gap which, as is explained later, is tripped in conjunction with the transient. The heating condenser is often included to maintain a hot low resistance spark, while the recharging supply arrives *via* high resistance from the d.-c. source which is suitably grounded at one end or at the mid-point.

For economical use of the oscillograph dimensions it is usual practice to distribute the total length of time sweep about the axis. This has usually been achieved by the use of magnets to bias the beam before it suffers the main deflection so that at zero time-sweep voltage it is well off the photographic plate. Burch and Whelpton[12] have pointed out the danger of some loss of focus of the beam with the strong magnetic field thus opposing the electrostatic one when the finish of tracing is a considerable way off the end of the recorder. They therefore use a time-sweep circuit which gives full deflection at the end of tracing and so minimizes the necessary strength of the electromagnetic bias which is, however, still required. The authors, on the other hand, have recently developed a method in which the connected circuit provides the required bias while yet observing the desirable equal division of deflection potential about earth. The system is applicable to bias control in the ordinate direction replacing the electromagnetic method here also, bias in this direction being necessary to superimpose the time calibration oscillation or to allow a large number of traces to be made on one record.

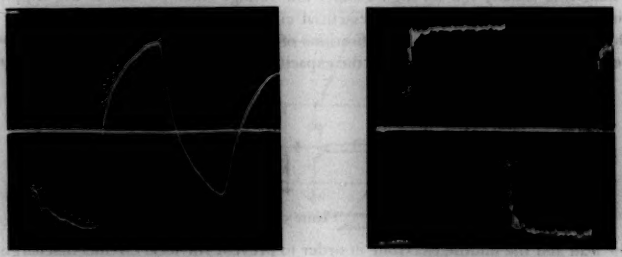
§ 7. TRIPPING OF THE OSCILLOGRAPH AND ITS SYNCHRONIZATION WITH THE TRANSIENT

The sphere of utility of the high-speed oscillograph embraces both work in the field and in the laboratory. In the former case with uncontrolled transients such as lightning, whose time of arrival cannot be anticipated, it is absolutely essential that the transient should itself initiate the oscillograph operation, and in general most oscillograph installations are laid out for transient tripping, even where the work is entirely in the laboratory. In the latter, however, where the transient may be under control, it is a matter of choice whether the transient should operate the oscillograph or whether synchronization of the two should be inherent in the local initiation of the transient.

In the first case, after the transient has tripped the oscillograph it is essential that its advent at the electrodes should be delayed by an amount corresponding to the small delay occurring before the oscillograph comes into operation. This is

universally obtained by interposing a suitable length of cable between the high voltage potentiometer and the oscillograph.

The use of a delay cable in conjunction with a potentiometer requires special design of its terminal connexions in order to avoid distortion and this aspect is considered in a later section. Apart, however, from the question of terminal distortion, an important factor is attenuation in the cable itself which becomes apparent as a smoothing off of the higher equivalent frequencies. This distortion tends to be excessive at high equivalent frequencies in the commercial type solid dielectric cables which are still widely employed for this purpose, the inherent distortion of any concentric cable here being aggravated by dielectric loss and the typical close ratio of conductor diameters. The inherent distortion can be reduced to a negligible amount for normal equivalent frequencies even in considerable lengths of cable, and



(*a*) (*b*)

Figure 4. Comparison of a rubber-insulated cable A with an air-insulated cable B.
(*a*) Reflected waves on cable A; (*b*) Reflected waves on cable B.

Reproduced by courtesy of the Institution of Electrical Engineers.

Allibone employs for this reason a suitably designed air dielectric cable, an oscillogram illustrating the comparative distortion in this and a normal solid dielectric cable being shown in figure 4.

On the other hand in the laboratory where the transient is under control the whole question of cable distortion in the measuring circuit can be eliminated by tripping the oscillograph a suitable short time interval in advance of the transient. In addition dispensing with the cable permits, especially in the case of the capacity type, the design of a potentiometer which imposes greatly reduced loading on the test object. Methods of synchronization inherent in the initiation of the transient are of considerable importance therefore and are frequent laboratory practice.

At any rate one American laboratory[56] employs this method, a locally generated impulse energizing the cathode supply and time sweeping circuit and after a suitable delay the high voltage generator. Certain German workers[8, 10, 50] have also described the use of impulse circuits initiated at sparkover of the common oscillograph auxiliary gap and feeding through a small transmission line to the test piece and

the oscillograph ordinate connexions, the line providing the necessary lag. The method, however, is restricted as regards the range of impulse testing possible. In one such method employed by the authors in the laboratory a small additional surge generator feeds a tripping impulse to the oscillograph and another to the main surge generator a fraction of a microsecond later.

For high voltage impulse work the synchronizing coupling or tripping mechanism is usually a critically adjusted spark gap which a small superimposed impulse causes to sparkover. Generally* too a three electrode gap is employed, the arrangement being amphipolar when the surge applied to the middle electrode causes sparkover to one of the outers followed by breakdown of the whole gap.

A tripping impulse of the order of a kilovolt, which for uncontrolled impulses is a fraction of the front of the transient obtained from a potentiometer or a pick-up antenna, is usually fed through a short low voltage shielded lead to the tripping circuit at the oscillograph. The essential circuit of the tripping arrangement is shown in figure 5 although modifications of this are variously adopted. Various devices such as high resistances or the capacitance as shown are interposed between

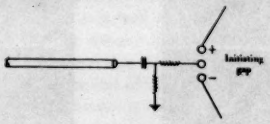


Figure 5.

the lead and the middle electrode in order to prevent the local circuits discharging back into the cable and to minimize the energy injected from the tripping impulse into the oscillograph circuits.

In special applications where the voltage available for tripping is small the gaps can be rendered more sensitive by employing magnesium spheres. This end is also furthered by substituting a smaller and more or less pointed electrode for the middle sphere. The practice of ultra-violet irradiation[32, 33] of the gaps is also frequently followed in order to obtain very consistent and rapid operation of these circuits with low trip voltages.

For initiating the oscillograph circuits valve tilting or "Kipp" relays are also employed[62, 63, 64, 65, 66]. Here a small impulse such as is picked up on an antenna suddenly lowers by means of a trigger action the impedance of a local valve circuit through which the oscillograph circuits discharge. The device is usually made amphipolar and has the advantage that only a very low operating voltage is required. However, it has the disadvantage of a greater time lag (of the order of 0·5 μsec.) than that of the three sphere spark gap and for this reason and because in high-voltage work its great sensitiveness is unnecessary, it does not find very wide application, although, where only very low tripping voltages are possible, its use is

* Certain German workers [8, 60, 61] have employed two sphere gaps tripped by an impulse applied through blocking condensers or coupling wires across the gap. The method appears to be much less sensitive than the three-sphere one.

more or less essential. In the three-phase medium frequency instrument of Whelpton, where the mechanical time motion and the electrical circuits are not interdependent, it is interesting to note that the de-trapping of the beam for a predetermined time is achieved by the use of a Kipp relay.

In concluding this discussion it is perhaps desirable to indicate in outline a few composite circuits typical of the various methods of tripping and synchronization discussed in the paper. Methods of testing where the impulse is synchronized with a crest of the power frequency pressure wave have recently come into prominence. This synchronizing can be effected by feeding an extra electrode in the first interstage gap of the surge generator with the required power frequency voltage and in the circuits of figures 6 and 9 it will be noted that synchronization is similarly provided in the auxiliary generators.

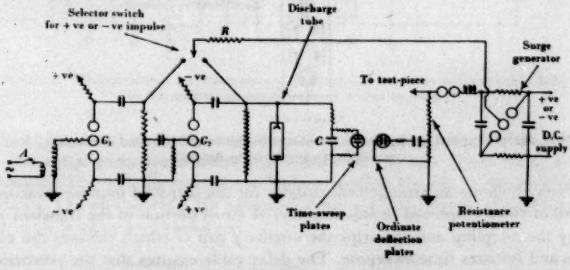


Figure 6. Typical American circuit for controlled surge measurement in the laboratory. Cathode energized only during recording.

Figure 6 shows an example of an American circuit for controlled impulse work in the laboratory. The method of cathode energization only during the time of photographing is, however, typical of many American circuits. Here closure of switch A energizes the mid-sphere of G_1 which trips on a 60-cycle peak and in turn trips G_2. This latter imposes a steady negative supply on the cathode of the discharge tube and also by charging up C provides the time sweep motion (note it is the charging action which is employed for this purpose). Either a positive or negative impulse is simultaneously fed through the delay resistance R to trip the Marx generator and so no cable delay is required.

Figure 7 is representative of American practice for uncontrolled impulses. In the upper circuit energization of the cathode and the time sweep are initiated from the trip potentiometer. This and the measuring potentiometer which is connected to the oscillograph directly without inclusion of a delay cable are generally in the form of insulator strings with additional rings. Thus no attempt is made to reproduce completely the front of the transient. Note too the double-sweep time-motion employed. When continuously excited discharge tubes are employed the

Norinder relay as shown in the lower sketch is used instead of the impulse connexion to the cathode.

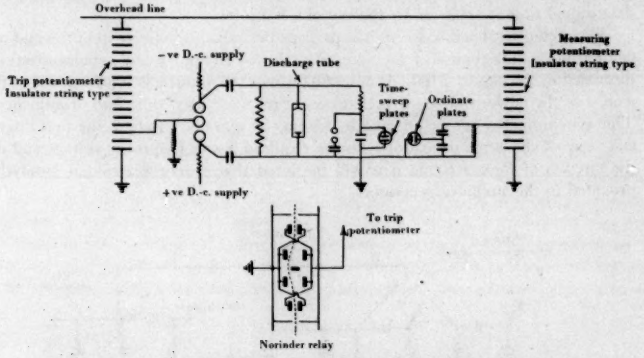


Figure 7. Typical arrangements for measurement of uncontrolled transmission line impulses, as employed in America.

Figure 8 shows an arrangement suitable for uncontrolled impulse measurement typical of those employed in this country. A small portion of the transient picked up by the coupling antenna trips the auxiliary gap *G* which releases the cathode beam and initiates time-sweeping. The delay cable ensures that the potentiometer

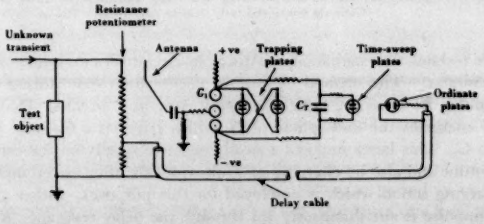


Figure 8. Typical circuit for measurement of uncontrolled impulses with continuous cathode energization used in this country.

reproduction of the transient arrives on the ordinate plates only when the oscillograph is in operation. Here continuous cathode energization is employed. The circuit is, of course, suitable for transmission line work by the substitution of a condenser divider for the resistance divider. Note the double plate trap system.

Figure 9 represents an arrangement as used in this country for controlled impulses. Closure of the press button *A* trips the auxiliary surge generator (if

necessary on a crest of the potential wave fed to the middle sphere) which trips the auxiliary circuits and a short interval later the main surge generator. Accordingly, the use of a delay cable connecting to the measuring potentiometer is obviated. Note also the suitability of this arrangement for measuring internal voltages in the test object.

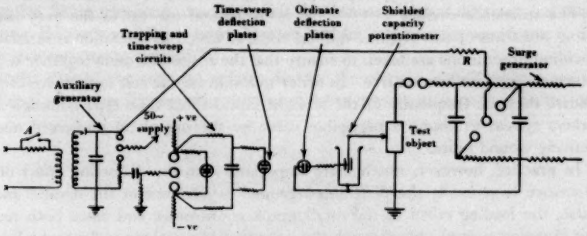


Figure 9. Typical circuit for measurement of controlled impulses without cable delay in the measuring circuit as used in this country.

Figure 10 shows an arrangement typical of German practice at low voltages. Here the surge generator delivers the impulse to the test object along a short transmission line and the transient voltage is applied directly to the oscillograph electrodes. The line provides the necessary delay while the beam release and time-sweep circuits come into operation.

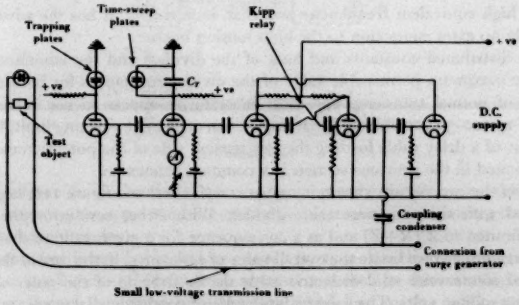


Figure 10. Low voltage testing arrangement typical of German practice.

In the particular arrangement shown the time-sweeping and trap circuits are tripped by means of an amphipolar valve tilting relay instead of by the usual three sphere gap method. Note that separate voltage supplies can be employed for these two circuits.

§8. VOLTAGE DIVIDERS

For high voltage impulse work two main types of voltage divider, the resistance and the capacitance, are available. An important virtue of either type apart from its faithfulness is that it shall impose a minimum loading on the connected circuit.

The resistance divider[12, 14, 67, 68, 69] is in general use up to the very highest testing pressures of two or three million volts, at which its operation is satisfactory providing precautions are taken to ensure that the resistance units together with all connected leads are corona free. In order to minimize the self inductance of these dividers they are frequently of the straight liquid filled tube type although some workers appear to obtain a negligible value by the use of oil immersed non-inductively wound units.

In practice, however, much more important than the distorting effect of self inductance is generally the distributed ground capacitance of the divider and, of course, the loading effect of the oscillograph connexions, and since both receive their charging current only through the potentiometer resistance they set a limit to its value for non-distorted reproduction. It is the practice of some workers to minimize the effect of the self capacitance by arranging the divider to lie as far as possible along the main high voltage field to earth. It has been pointed out, however, that inductive fields linking the divider may themselves give rise to distortion.

It is sometimes the practice* to dispense with the use of a separate resistance divider and employ instead a suitable tapping on the surge generator discharge resistance with the oscillograph connexion made through a small capacitance and a short shielded lead. Its sponsors state that it gives faithful reproduction for relatively high equivalent frequencies and that, incidentally, it has the advantage of requiring no extra connexion to the high tension busbar.

The distributed constants and bulk of the dividers and the laboratory layout limit the maximum permissible value of the divider resistances for faithful reproduction of normal transients and most laboratories appear to use values of the order of 15,000–30,000 ohms.† What imposes an even more exacting limit, however, is the use of a delay cable loading the low tension side of the potentiometer which as mentioned in the previous section is a common connexion.

One of the two circuits shown in figure 11*a* (Gábor) and figure 11*b*, is generally employed with the resistance cable divider. With either connexion the divider ratio is limited to $Z/(R+Z)$ and as a consequence for a given ratio of division the cable surge impedance limits the overall value of resistance, in the case of the widely employed commercial solid dielectric cable the limit being of the order of 20,000 ohms for a million volts. The use of a large ratio of conductor diameters and an air‡ dielectric as followed by Allibone[14] gives a surge impedance of as much as six-fold higher than the commercial cable and permits a corresponding desirable increase

* For example, see [68].

† Very recently The National Physical Laboratory have found it necessary to employ values as low as 2000 ohms.

‡ The National Physical Laboratory also employ an air dielectric cable[20].

of the limiting resistance, though for a given purpose the limit set by the stray capacitance still stands.

So far as the range of usefulness of the resistance divider is concerned, with their usual practice the Americans[70] state that faithful reproduction at very high voltages is given of wave fronts as fast as 0·5 μsec. while they do not recommend its use for faster transients than this. In this country with a different technique Allibone, Edwards and Perry have obtained faithful response for wavefronts as fast as 0·2 microsecond at 750 kV.[20]

In the case of the capacity voltage divider the obvious considerations are the avoidance of corona (if anything this being more important here than in the case

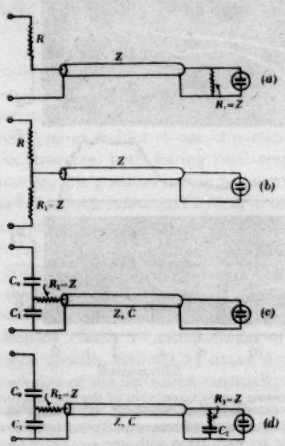


Figure 11.

of the resistance divider) and the necessity for preventing, either by screening or by suitably disposing the apparatus, any extraneous electrostatic field linking the middle or measuring electrode.

The simplest form of this divider as used in some American laboratories[69, 70], employs a small pick-up electrode or antenna mounted in the main high voltage field to earth and connected by a short screened lead to the oscillograph. The device of course requires careful layout, is only suited to measure terminal potentials to ground, and to maintain a constant calibration requires an unaltered geometry of surrounding apparatus. For work on energized transmission systems line type insulators fitted with three extended rings are fairly generally employed in America[51, 52]. By construction not being inherently shielded they are preferably mounted on the ground some distance from the line and require to be axially

disposed along the field of their h.-t. lead to earth. By reason of their small overall capacitance both the above types of divider are unsuited to delay cable loading.

The ring type potentiometer where the middle ring is of smaller diameter than the outers and is fairly completely shielded by the earth ring does not require such careful positioning and has been employed by the authors up to the highest voltages for measurements not only of terminal waves but of internal voltages in apparatus.

The concentric cylinder capacitive divider in the form of the condenser bushing type has been frequently employed for testing and transmission line work and having characteristically a greater capacitance it can be suitably loaded by a delay cable.

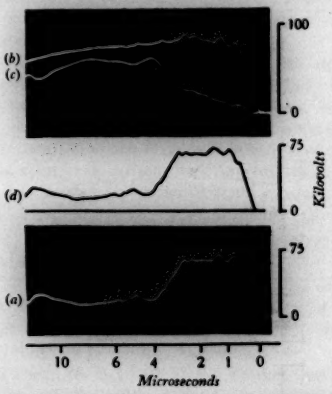


Figure 12. (*a*) Voltage appearing across 10 per cent of the line end turns of a transformer winding. (*b*) Voltage to earth at the transformer terminal. (*c*) Voltage to earth at a point 10 per cent down the winding. (*d*) Arithmetic difference of (*b*) and (*c*); (*a*) was taken with two condenser potentiometers; (*b*) and (*c*) were taken with one condenser potentiometer. (Reproduced by courtesy of The Institution of Electrical Engineers.)

Specifically designed for the purpose[20] with the measuring electrode totally enclosed the concentric cylinder arrangement can be completely shielded and although it suffers from the general disadvantage of the condenser type in respect of bulk at high voltages its capacity may be designed as required and its field of application is less restricted than that of any other divider. Two balanced dividers of this last type have been employed by the authors up to 300 kV. for the measurement of internal voltage gradients in transformer windings and despite the fact that such a winding is very sensitive to loading good accuracy as is illustrated in figure 12 is obtained. Norinder[67] has employed capacitive dividers for the same purpose, but in spite of its engineering importance this measurement does not appear to have been described by others.

One happy characteristic of the capacity type divider is the very low value of

loading of a few micromicrofarads only which they impose (and which properly decreases with increasing voltage range). Unfortunately, when used with a delay cable it is just this feature which is jeopardized, the value of the overall capacitance requiring to be increased to limit distortion.

The simplest connexion, used by Gábor, for the capacity cable divider is that shown in figure 11*c*. Here the imperfect absorption of reflections by R_s requires that the capacitance C_1, should be large compared with the cable capacitance C in order to avoid distortion, this also requiring an increased overall capacitance for a given ratio. The circuit of figure 11*d* introduced by F. P. Burch[71] enables this capacitance to be reduced sixfold for the same conditions of ratio and distortion and so represents a great advance on the simpler one, but the overall capacitance is nevertheless still necessarily increased in proportion to the cable length for a given permissible distortion.

Other methods of employing delay cables in conjunction with both resistance and capacity potentiometers are employed, but those described are the more general and important and a detailed discussion is not possible here.

The question of which type of divider to use in particular applications is still very much a matter of controversy, both having their respective limitations and drawbacks. As the authors see the position it may be summed up thus.

The resistance divider being little influenced by length of cable may impose least loading where delay cables are required and although inherently limited as regards frequency response with careful design the limit may be higher than is required for many researches. A point not usually mentioned is that although it is less bulky than the corresponding shielded capacity type it requires to be accorded about the same space. It cannot be used for work on energized apparatus and its unshielded nature restricts its application chiefly to measurement of terminal voltage. An advantage of the resistance divider, claimed by many American engineers[70], is the fact that, with a knowledge of the deflection sensitivity* of the oscillograph, it provides a fundamental standard for measuring voltage.

The shielded capacitance potentiometer imposes less loading on the high voltage circuit, at any rate when no delay cable is used, and in addition that loading is of the same nature as the incidental capacitance to ground. These two facts, together with the effective shielding obtainable, render this divider definitely the more suitable for measurements of internal voltages. In any case its response to steep wave fronts is not limited as in the case of the resistance divider and, of course, either this or a hybrid resistance-condenser divider is of necessity employed on energized apparatus. A drawback is its comparative bulk and expense for high voltages. Compressed gas, solid dielectric and oil-filled capacity potentiometers effect a considerable saving in bulk but in this connection the general question of loss in non-gaseous dielectrics has not yet been explored.

The general acceptance of the limitations of the resistance potentiometer as regards frequency response and the dissatisfaction with the capacitive potentiometer

* Finch determines the accelerating voltage by diffracting the beam by passage through a thin gold film.

expressed by some authorities have been considered reasons for dispensing altogether with dividers where possible by employing high voltages direct on the deflection system. While on principle the elimination of any suspected apparatus is obviously desirable the authors are of the opinion that a shielded air-dielectric capacity divider with normal deflection system does not distort to any practical extent more than does a high voltage deflection electrode design under the same conditions.

Some time ago Binder[28] constructed an instrument suitable for the direct connexion of pressures up to 200 kV. and recently Nuttall[20] has developed an oscillograph suitable for the application of pressures up to 100 kV. to ground directly on either deflection plate. While it has been constructed primarily for very rapid work such as the breakdown of spark gaps on the front of very steep waves, this instrument may also give an interesting comparison of results with similar ones obtained employing capacitive dividers.

With all potentiometer arrangements it is of high importance that a low resistance earth connexion should be employed and many workers have investigated and stressed the importance of solidly bonding the potentiometer earth or cable sheath to the oscillograph casing and the earthed ordinate deflection plate. Where good earthing is not available some American workers[68, 69] have stated that they balance out the local earth voltages by connecting the "live earth" to the other ordinate deflection electrode, but the practice can only be regarded as an expedient, and for accurate measurement of rapid phenomena at high voltages very careful layout of a large capacity earthing system about the oscillograph is necessary.

§9. THE MEASUREMENT OF SURGE CURRENT

The measurement has been considered so far throughout this paper of voltage transients only, and while it is in this sphere that the high-speed oscillograph has mainly been employed, no discourse on the subject of high-speed oscillography would be complete without reference to the measurement of current transients.

This aspect of the subject has come into prominence owing to the recent need to simulate and measure surge currents in the laboratory of magnitudes similar to those found in lightning strokes and in transmission line towers when struck by lightning. Stroke currents have been measured and estimated by several independent observers chiefly by means of the surge crest ammeter and are found to range between about ten thousand and a few hundred thousand amperes[72, 73].

The information so far available regarding the use and measurement in the laboratory of currents of such extreme magnitude persisting for times only of the order of microseconds comes almost entirely from the United States[74, 75]. The two main problems involved are the arrangement of the surge generator for a very low value of circuit inductance and the design of low resistance non-inductive shunts.

The first is outside the scope of oscillography proper. So far as the shunts are concerned, the important considerations of course are very low inductance,

negligible skin effect, and independence of resistance on temperature, and one arrangement which has proved very satisfactory for currents up to 120,000 amperes comprises a score or so of 20 s.w.g. manganin wires bent double, twisted together, paralleled and arranged radially on an insulating disc. Bellaschi, for instance, has shown oscillograms obtained using such a shunt of current pulses of this order involving wavefronts less than a microsecond.

§ 10. CONCLUSION

The paper purports to describe the position of high-speed oscillography at the end of 1934, and surveys the greater part of the literature which has appeared during the preceding twelve years. While there has been every intention of giving full weight to all published information which the authors from their own practical experience have adjudged relevant and of value, it is possible that some important publications have been incompletely apprehended or inadvertently overlooked. It is sincerely hoped that such transgressions have been few in number.

The bibliography on this subject contains a considerably greater number of papers than is quoted here, but for the sake of conciseness and for the reason that many authors (particularly German ones) superseded their own work by an almost immediate second publication, similar papers by the same writer have not all been included, except where questions of priority appeared to be at stake.

It is well to state in this conclusion, which has been written separately at the beginning of 1936, that although the paper is intended to summarize the situation at the end of 1934, this has not been materially altered by information which has subsequently appeared. It is thus evident that apart from the very important subjects of high-voltage and high-current potentiometry, and the development of lower voltage oscillographs with relatively high writing speeds, a state of temporary quiescence has been reached. No doubt the various workers and laboratories will continue for some time, with minor developments, along the individual lines indicated in the paper. *Adhuc sub judice lis est.*

§ 11. ACKNOWLEDGMENT

The authors gratefully acknowledge their appreciation of the interest shown by Ferranti Ltd. in the production of this paper.

PART 3. THE ELECTRON MICROSCOPE

By H. J. H. STARKS, Ph.D., D.I.C., A.Inst.P.

§ 1. PRINCIPLES OF ELECTRON OPTICS

In 1926 Busch[1] showed that an analogy existed between geometrical optics and the behaviour of electrons in magnetic and electric fields. Later workers have developed these analogies more completely and have shown that such fields possess the properties of a general optical system.

(*a*) *Magnetic focusing.* The earliest form of magnetic focusing of an electron beam is that given by a homogeneous magnetic field produced by means of a long solenoid. The beam is coaxial with, and wholly within, the solenoid. Electrons in motion along the axis of the magnetic field are unaffected by the field, whilst the projection of the paths of those slightly inclined to the axis on a plane perpendicular to it are circles; the resultant paths of the electrons are thus helices.

If an electron of mass m and charge e travels with a velocity v at a small angle α to the direction of a uniform magnetic field H, then the radius r of the circular orbit is given by

$$r=\frac{mv\sin\alpha}{eH}.$$

The time t required for the electron to traverse the orbit (i.e. one turn of the helix) is given by

$$t=2\pi m/eH.$$

It follows that the time is independent of v, r and α and thus all electrons leaving a given point B on the axis will meet again on the axis at a certain point B', distant l from B, simultaneously and with equal longitudinal velocity.

Since $$l=\frac{2\pi v\cos\alpha}{He/m}, \qquad (\text{i.e. } l\propto 1/H)$$

a definite value of H is required to focus the electrons at a desired place. If the electron source is a point then the electrons are focused to a point but if the source has a finite area which is small then an upright image of the same size is obtained in the case of paraxial electrons. Use has recently been made of this fact in the electronic television system due to Farnsworth[2].

Busch[1] deduced the above relations as a particular case of the general problem of the behaviour of electrons in axially symmetrical electromagnetic fields. He considered also the case of a magnetic field of limited range of action produced by means of a short solenoid; he showed that this behaved like a convex lens and that the conjugate distance relation

$$1/l'-1/l=1/f' \qquad (f'=\text{second focal length}),$$

held if the initial and final energy potential of the electron was the same and if the conjugate distances l and l' of the object and image respectively were long compared with the effective range of the magnetic field. Provided object and image are in a region free from field, the conjugate distance relation should hold whatever the effective range of the field. For a "thin" lens

$$1/f'=\eta/8V\int_{-\infty}^{+\infty}H_z^2\,dz,$$

where η is e/m, V is the energy of the electron traversing the field, and H_z is the axial magnetic field strength at any point z. Integration should only be performed over the effective range of the field, but since the field has a negligibly small value outside this range, we can integrate between infinite limits. It can be shown that,

for a short solenoid of length $2l$ and radius r, wound with n turns per centimetre, and carrying a current J e.m.u.,

$$f' = \frac{16Vr}{3\pi^2\eta n^2 J^2}\{1 - \tfrac{8}{9}(l/r)^2\} \quad \text{approximately.}$$

These formulae hold only for paraxial conditions and on the assumption that the initial velocity is the same for all the electrons in the beam.

A special property of the short solenoid magnetic lens as compared with the long solenoid (uniform field) lens is that a paraxial beam in a plane containing the axis will, after refraction by the lens, lie in a different plane which still contains the axis but which is rotated relative to the first plane through an angle ϕ, where

$$\phi = -\frac{1}{2\sqrt{2}} \cdot \sqrt{\eta} \int_{-\infty}^{+\infty} \frac{H_z}{\sqrt{V_z}}\, dz \quad \text{approximately,}$$

H_z being the axial magnetic field at the point z and V_z the energy of the electron at that point.

(*b*) *Electrostatic focusing.* An electron, moving under the action of an electric field V and incident on an electrostatic field layer (i.e. two plane surfaces close to one another and at potentials of V and $V+V_1$), obeys the same laws as a beam of light incident on a surface separating two media of different refractive index (tangential velocity remains constant). To a first approximation,

$$\sqrt{V} . \sin\theta = \sqrt{(V+V_1)} \sin\theta',$$

where θ and θ' are the angles of incidence and refraction respectively. This expression is analogous to the ordinary optical equation

$$n \sin\theta = n' \sin\theta',$$

where n and n' are the refractive indices of the media containing the incident and refracted rays respectively. The relative refractive index n of the layer of electrostatic field may thus be written in the form

$$n = \sqrt{(1 + V_1/V)}.$$

V_1 may have any value, positive or negative, with respect to V and thus may n vary between wide limits. If V_1 is positive (i.e. if V_1 has an accelerating action on the electron) n is greater than 1 but if V_1 is negative (i.e. if V_1 has a retarding action on the electron) n is less than 1.

Refraction occurs at a spherical potential layer of radius r, as in the case of optical refraction at a curved surface. Assuming paraxial conditions, it may be shown that conjugate points exist on an axial line containing the centre of the layer and that

$$\frac{\sqrt{V+V_1}}{l'} - \frac{\sqrt{V}}{l} = \frac{\sqrt{V+V_1} - \sqrt{V}}{r},$$

where l and l' are the object and image distances respectively and the usual Cartesian sign convention is used. By making l' infinite and V_1 small in comparison with V, it is clear that $l = -2rV/V_1$ to a first approximation; this corresponds to the "first focal length".

If a symmetrically disposed refracting surface is added to the former, the analogy of a simple lens is obtained. In general, the refraction of an electron in an electric field is continuous and the equivalent lens has an infinite number of refracting surfaces. The field may also be likened to an optical system consisting of an axially symmetrical variation of refractive index. It can be shown that, for a "thin" electrostatic lens, the conjugate distance relation is

$$\sqrt{V'}/l' - \sqrt{V}/l = \frac{1}{8}\int_{-\infty}^{+\infty} V_z^{-\frac{3}{2}} \left(\frac{\partial V_z}{\partial z}\right)^2 dz,$$

where V and V' are the initial and final energy potentials of the electron and V_z is the energy of the electron at the point z. If, in addition, a magnetic field is superposed on the electric field, the relation is

$$\sqrt{V'}/l' - \sqrt{V}/l = \sqrt{V}/f',$$

where
$$1/f' = \frac{1}{8\sqrt{V'}}\left\{\int_{-\infty}^{+\infty} V_z^{-\frac{3}{2}} \left(\frac{\partial V_z}{\partial z}\right)^2 dz + \eta \int_{-\infty}^{+\infty} V_z^{-\frac{1}{2}} H_z^2 \, dz\right\}.$$

The focusing action of an electrostatic lens may be expressed in terms of the curvature of the equipotential surfaces on either side of the lens, remembering that the electron is accelerated in the direction of increasing potential, but that if the velocity be increased, the deviation produced by the lens is diminished. A symmetrical lens, i.e. one which has similar equipotential surfaces on either side, has been termed a "single" lens (*Einzellinse*). A lens which is unsymmetrical (i.e. one in which the initial and final potential energies are different) has been termed an "immersion" lens.

(*c*) *Space-charge focusing.* It has been known since about 1900 that residual gas causes a concentration of an electron beam. A beam which is divergent in a high vacuum may become convergent if a trace of gas is introduced. This converging action has been explained by Johnson[3] in terms of the difference in mobility of the electrons and positive ions. Some of the electrons in the beam collide with gas molecules and ionize them, but owing to the relatively large mass of the ions and their smaller mobility, comparatively few drift out of the beam and thus positively charged ions accumulate along the beam. The electrons which drift out of the beam set up a region of negative electrification round it; the resulting space-charge field is approximately an axially symmetrical one and it may be used for focusing. The electron optical properties of the system are, however, largely dependent on the nature and pressure of the residual gas, a fact which is not conducive to perfect imagery, but which, in certain cases (e.g. in the usual form of cathode-ray oscillograph) may be of practical value.

§ 2. ELECTRON-OPTICAL ABERRATIONS

The discussion hitherto has been confined to the case of paraxial conditions. For extra-axial electrons a number of aberrations are observed analogous to those met with in the geometrical optics of extra-axial pencils. A variation in focusing distance with aperture is found (this is analogous to spherical aberration) and will

give rise to aberrations of the comatic and astigmatic type. The focusing action of electrostatic or magnetic lenses depends on the primary velocity of the electrons, so that an effect analogous to chromatic aberration will be produced if the electron beam is not monochromatic. Thus in electric fields the following aberrations are present; spherical and chromatic aberration, coma, astigmatism, curvature of field and distortion. Similar aberrations exist in magnetic fields; in addition there is the aberration due to the rotation of the image and to the variation in rotation with aperture and image size. Martin has termed the differential rotation of the image which would be present with an infinitely small aperture "image twisting" and the differential rotation when the aperture is increased "image wringing". Thus in a magnetic field there are three additional aberrations, viz. rotation, which in no way impairs the image qualities, "image twisting" which causes tangential displacement of the extra-axial image points, and "image wringing" which leads to a loss in definition of those points.

§ 3. BRIEF SUMMARY OF WORK ON ELECTRON OPTICS

Theoretical work. A number of purely theoretical papers on electron optics have been published since that of Busch. Scherzer[4] has calculated the focal length of a short electrostatic lens given the potential distribution along the axis of symmetry and has deduced the primary spherical aberration. Picht[5] has confirmed these calculations and has derived the positions of the principal planes and nodal points. Posener[6] has investigated the focusing of extra-axial beams, Ruska[7] has discussed methods of focusing beams of large cross section, and Ollendorf and Wendt[8] have applied relativity corrections to Busch's calculations (this correction becomes negligible for voltages below 10 kV.). Glaser[9] has worked out in great detail the electron optics of axially symmetrical electric and magnetic fields and has developed fully the analogy between geometrical optics and electron optics.

Experimental work: electrostatic focusing systems. Knoll and Ruska[10] have described the electron-optical behaviour of a number of electrostatic lenses constructed of wire gauze nets to which potentials of 3–30 kV. were applied. Images of a small electron source and of a series of holes in a diaphragm were obtained and the relation between the focal length and the potential distribution of the lens was calculated from Busch's equations and confirmed by experiment. Brüche[11] and Brüche and Johannson[12] have used wire net electrostatic lenses to focus the electrons emitted by an oxide cathode on a fluorescent screen using low voltage electrons (100–650 V.); whilst Johannson and Scherzer[13] have described a low voltage electrostatic lens and verified the conjugate distance relation. A large number of workers engaged in research on cathode-ray oscillographs have also described various types of electrostatic lens. Amongst these workers may be mentioned Hamacher[14], Graupner[15], Brüche[16], Zworykin[17] and Ruska and Borries[18]. Johannson[19] has introduced his "immersion objective", consisting of two or more diaphragms placed a short distance apart and at suitable potentials. With a four electrode lens he has obtained focal lengths of the order of 1 mm.

Experimental work: magnetic focusing systems. Short magnet coils have been used since about 1903 for converging electron beams but it was only after the pioneer work of Busch in 1926 that their value as magnetic lenses was fully appreciated. Knoll and Ruska[20] have constructed an apparatus with which the influence of one or more magnetic lenses on a beam of high speed electrons (30–80 kV.) has been examined. The optical constants of the lenses have been derived from Busch's calculations and verified experimentally. Knoll, Houtermans and Schulze[21] have described an electron microscope in which two magnetic lenses have been used to image an oxide cathode on a fluorescent screen using electrons of 100–2000 V. Ruska[22] has described a high-power magnetic lens for use with 75 kV. electrons. Two iron-clad magnet coils with pole pieces separated by a small distance have been constructed so that the effective magnetic field has been confined to a narrow region between the pole pieces. Focal lengths of about 2 mm. have been obtained in this way by Ruska.

Experimental work: combined focusing systems. A number of workers have combined an electrostatic with a magnetic method of focusing in an endeavour to produce the analogue of an achromatic combination. Knoll and Ruska[10], Johannson and Knecht[23], Knecht[24] and v. Ardenne[25] have described combination methods in which the electric lens is used as a diverging lens and the magnetic as a converging lens. Glaser[9] showed that a "thin" purely magnetic lens was only capable of acting as a converging lens but that, if an electric field (e.g. that due to space-charges) existed along the axis of the magnetic field, the system may act as a diverging lens. A "thick" magnetic lens may be convergent or divergent (in ordinary optics a system containing a number of components of positive power may be equivalent to one of negative power).

The expression for the focal length of a magnetic lens has been given above. In the case of an electrostatic lens the corresponding expression may be written

$$1/f' = -\frac{1}{4\sqrt{V'}}\int_{-\infty}^{+\infty} V_z^{-\frac{1}{2}}\left(\frac{\partial E_z}{\partial z}\right) dz,$$

where V_z is the potential at a point z on the axis and E_z is the axial field at the point z. Since the focal lengths of thin low-power electric and magnetic lenses tend to be proportional to the energy potentials of the electrons, it will be evident that the achromatism of thin lens doublets in the optical sense may not be readily attainable. The theory of this question does not seem to have been adequately examined.

§ 4. ELECTRON MICROSCOPY

The principles of electron optics have, up to the present, found their most important applications in the design of the cathode-ray oscillograph, the electron diffraction camera and the electron microscope.

The potentialities of electron microscopy first seem to have been realized by a number of German workers, notably Brüche, Johannson, Knoll, Ruska and Richter.

The most attractive feature lies in the greatly improved resolving limit (R.L.) which is theoretically obtainable, as compared with light.

An estimate of the R.L. of an electron optical lens may be made if the numerical aperture (N.A.) of the lens is known and if it is assumed that the electrons constituting the beam have a wave-length given by

$$\lambda = \sqrt{(150/P)} \times 10^{-8} \text{ cm.},$$

where P is the potential drop in volts equivalent to the energy of the electron (see G. P. Thomson, *Wave Mechanics of Free Electrons*, p. 47). In geometrical optics the R.L. of a system is given by $0{\cdot}5\lambda/\text{N.A.}$ and if this expression may be applied to an electron optical system with a N.A. of 0·1 and 50 kV. electrons, the R.L. is $0{\cdot}27 \times 10^{-4}\mu$. ($\mu. = 10^{-3}$ mm.) (for a high-power microscope objective of N.A. 1·40 the R.P. is $0{\cdot}2\mu$. for sodium light).

At present the largest aperture which has been recorded for a high-power magnetic lens is 1/12 and this has given a useful magnification of about 8000 (in two stages), with a limit of resolution of about $0{\cdot}05\mu$. In ordinary microscopy the limit of resolution is about $0{\cdot}2\mu$., so that, even at this early stage in its development, the electron optical system has a R.L. exceeding that of a high-power microscope objective.

Brüche[11] and Brüche and Johannson[12, 26] have given a brief description of the application of electrostatic lenses to electron microscopy. Knoll and Ruska[20], Knoll, Houtermans and Schulze[21] and Richter[27] have constructed electron microscopes with magnetic lenses and have used them to form images of an oxide cathode. The electrons emitted by the cathode have been focused by the magnetic lens or lenses on a fluorescent screen and the distribution of emission from the cathode has been examined by visual and photographic means and its variation with temperature, time and gas concentration has been studied. Similar work has been carried out by Brüche and Johannson[26, 28] using an electron microscope with electrostatic lenses.

A typical form of electron microscope employing magnetic focusing is that shown in figure 1, which is due to Knoll and Ruska. A cold cathode discharge serves as a source of electrons which are accelerated by a potential of 10–100 kV. applied to the anode. The electron beam is concentrated by a magnetic lens (condenser coil); it then passes through the object and two further magnetic lenses which act as objective and projector lenses. These project an image of the object on to a fluorescent screen or photographic plate. The whole of the apparatus with the exception of the cathode chamber is highly evacuated.

Soon after it was realized that electrons derived from any source could be focused by a suitable electron optical lens a number of papers were published in which electron microscopes were described utilizing electrons derived from secondary and photoelectric emission, from metallic surfaces by reflection and from transmission through thin foil.

The secondary emission from a cold surface bombarded by electrons has been examined with an electrostatic electron microscope by Zworykin[17] and Knoll and Lubszynski[29]; Brüche[30] has studied the distribution of the photoelectric emission from a zinc plate under the action of ultra-violet light by focusing the photo-

electrons with a magnetic lens. Primary reflected electrons have been used by Ruska[31] to produce images of metals of high atomic weight, whilst Ruska and Borries[32] have used electrons transmitted through thin metal foils to reveal the internal structure.

The magnetic electron microscope has also been used by Brüche and Knecht[33] to investigate the crystallographic changes which occur in iron and nickel as these pass through their transformation temperatures; the transition from the α to the γ state in iron has been followed visually. An iron or nickel wire has a layer of barium deposited on it and, after suitable activation of this layer, an image may be obtained of the micro-crystal structure of the surface of the metal wire by focusing the electrons emitted by the active barium.

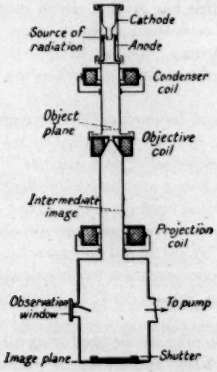


Figure 1.

The application of electron microscopy to the study of biological objects has been initiated by Marton[34]. An object for examination by electron optical means must be capable of withstanding the effects of a high vacuum and electron bombardment and must be so mounted as to minimize charging-up. Marton has impregnated cell structures (such as leaves and plants) with osmium in such a way that a framework of the object is preserved although the object itself is destroyed. This framework is then mounted on a copper net.

Recently attention has been directed to the construction of high power magnetic lenses and Ruska[22,35] has designed lenses which give a useful magnification of about 100, so that a double stage gives a magnification of about 10,000. Before a further advance is made in the design of still higher power magnetic objectives much work remains to be done on the aberrations of the lenses in an endeavour to increase the useful N.A. (which at present is of the order of 1/12).

To summarize briefly, the electron microscope has up to the present been applied to a study of the electronic emission from oxide cathodes, tungsten and thoriated tungsten filaments, vapour deposited and photo-sensitive surfaces, secondary emission from bodies bombarded by high-speed electrons, the macro-structure of thin films by transmitted electrons and the micro-crystalline structure at different temperatures of the surfaces of metals.

A comprehensive treatise on electron optics has recently appeared in German by Brüche and Scherzer[36] and reference may be made to this book for full details of the various applications of electron optics.

Part 4. THE ELECTRON-DIFFRACTION CAMERA

By G. I. FINCH, Imperial College of Science and Technology

An electron-diffraction camera consists of a device for producing a fine beam of electrons moving at a controlled velocity, an arrangement for holding the film or surface to be examined in any desired position relative to the beam, together with visual and photographic means for observing and recording the angular distribution of the diffracted electrons.

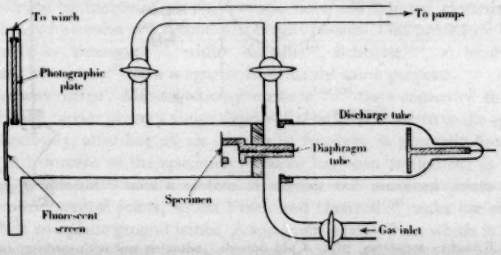


Figure 1. The first electron-diffraction camera; Thomson[1]. Cold cathode: induction coil as h.-t. source: collimation by tubular diaphragm: no specimen adjustment: camera length 30 cm.: exposure time 10 sec.

Since Thomson's[1] first camera (figure 1) was described in 1928 many others have been constructed more or less fulfilling the above requirements. The chief modifications in their design may be summed up in the following alternatives: (i) hot or cold cathode as electron source, (ii) production of a monochromatic electron beam by means of either (*a*) a constant accelerating potential or (*b*) magnetic or electrostatic filtration of polychromatic cathode rays, (iii) collimation of the beam by (*a*) a double or tubular diaphragm or (*b*) a single diaphragm followed by magnetic focusing, and (iv) specimen carrier construction capable of imparting to the specimen all six or frequently fewer degrees of freedom.

The production of the cathode beam. The cold cathode method of generating the electron beam in a gas discharge tube, used by Thomson[1, 3] (figure 1) and by Kikuchi[4] (figure 2) in the early electron diffraction cameras, has since been adopted by Wierl[7], Kirchner[9], Trillat and Hirsch[13], Finch and Quarrell[23] (figure 3) and others[2, 24, 26, 27] with satisfactory results whilst Muto and Yamaguti[15],

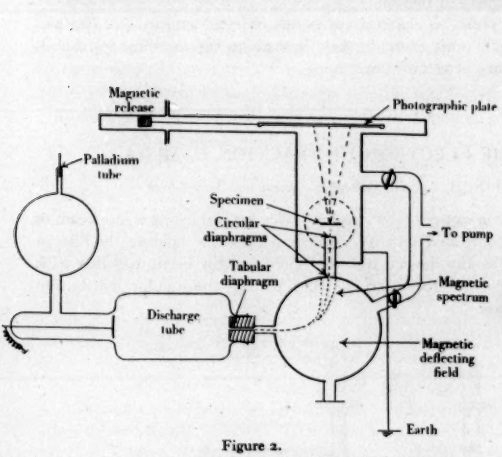


Figure 2.

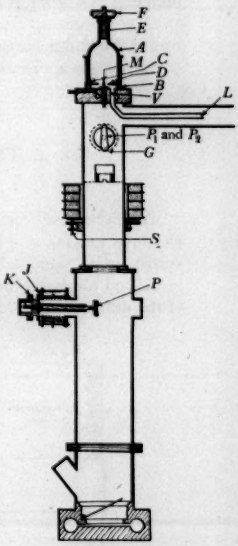


Figure 3.

Figure 2. Kikuchi's apparatus, 1928. Cold cathode: induction coil with mercury interrupter for h.-t.: monochromatization by magnetic filtration of polychromatic rays: collimation by multiple diaphragms: specimen given two rotational degrees of freedom: camera length 15 cm.: exposure time 20–120 sec.

Figure 3. Electron-diffraction camera; Finch and Quarrell[23]. Cold cathode: constant current h.-t. supply: collimation by single diaphragm and magnetic focusing: specimen has 6 degrees of freedom: camera length 50 cm. for polycrystalline specimens and 20 cm. for single crystals: exposure time 1/5 sec.

Trillat and Hirsch[13], Kirchner[14], Thomson[30], Ponte[8], W. L. Bragg and Darbyshire[16], Eisenhut and Kaupp[12] have employed hot filament generation in a highly evacuated tube. Further description of the different types of discharge tube is unnecessary as these do not differ materially from those in current use in high-speed cathode-ray oscillography. Finch and Quarrell[23, 25] have pointed out the advantages of the cold cathode cathode discharge tube. The range over which the cold cathode

tube can be used is approximately 15–75 kV. Below 15 kV. the electron beam contains too large a proportion of electrons which have lost energy through collision with gas molecules in the discharge chamber; on the other hand, it is difficult to obtain steady discharge conditions above 75 kV. It so happens, however, that the most useful range of accelerating voltages for structure analysis lies between about 30 and 70 kV. Under these conditions a cold cathode source is to be preferred to a thermionic emitter, not only for reasons of robustness and simplicity, but also because advantage may be taken of its peculiar properties to obtain a monochromatic electron beam. Of the three discharge conditions in the cold cathode tube, namely gas pressure, current and voltage, the fixing of any two determines the other. Since the pressure and current can be independently regulated, the accelerating voltage may be exactly controlled. The modern tendency undoubtedly favours the use of the cold cathode tube both in electron diffraction and high-speed cathode-ray oscillography.

The production of a monochromatic beam. Thomson[1, 3] and most other workers use a moderately well-smoothed high-tension source and collimate with a tubular or coaxial diaphragm system. Finch and Quarrell[23, 35] on the other hand, employ a constant-current high-tension source in conjunction with a controlled leak to the cathode chamber and monochromatize the beam passing through a single diaphragm by magnetic focusing. In 1928 Reid[2] deflected his polychromatic electron beam into a spectrum by means of an electrostatic field and selected electrons of the desired velocity by means of a second diaphragm system. This procedure has since been adopted by Yamaguti[37], whilst Kikuchi[4], Schobitz[10], v. Meibom and Rupp[15] and Mongan[19] use a magnetic field for the same purpose.

The specimen carrier. Although many workers[29, 36] have contented themselves with a specimen carrier giving a single degree of rotational freedom to the specimen, a greater flexibility, affording all six degrees of freedom, is generally found to be desirable. Adjustment of the specimen *in vacuo* has been carried out in different ways; thus, Thomson[3] uses a system of toggles and universal joints operated through ground conical joints, whilst Finch and Quarrell[35] make use of flexible metal bellows and plane ground joints. A specimen carrier[35] on which is mounted a reflection specimen is shown in figure 4, from which it will be evident how, among others, the following independent motions of the specimen may be obtained: (i) inclination of the specimen plane to the electron beam, (ii) translation into the beam, (iii) rotation in the azimuthal plane, and (iv) lateral motion in the specimen plane at right angles to the electron beam. Several types of specimen holders, each suited to a particular form of specimen, are interchangeable and screw into the main stem of the specimen carrier.

In work on vapours Wierl[7] and later workers[20, 24, 26, 41] use specimen sections in which a molecular beam of the vapour traverses the electron beam and is condensed on a liquid-air-cooled surface. Organic substances which are solid at ordinary temperatures are vaporized in a suitable oven.

The recording section. Visual observation of the diffraction pattern is effected by means of a fluorescent screen mounted on glass[1, 3] or metal[23], and photographic

recording is usually achieved by direct impact of the high-speed electrons upon a photographic plate *in vacuo*.

Thomson[1] mounts his fluorescent screen on glass so that rough measurements of the pattern may be made directly and lowers the photographic plate in front of the screen to obtain a permanent record. By suitable masking several independent exposures may be made upon the same plate. In the cameras described by Finch and Quarrell[23, 35] the photographic plate is stationary and is exposed by raising a metal shutter which is backed with fluorescent material, thus enabling observation to be made through a conveniently placed window.

The use of roll-film has been advocated by some workers[20, 38], but this introduces difficulties owing to the contraction on development and the amount of outgassing required. Indeed, the speed of operation of the modern electron

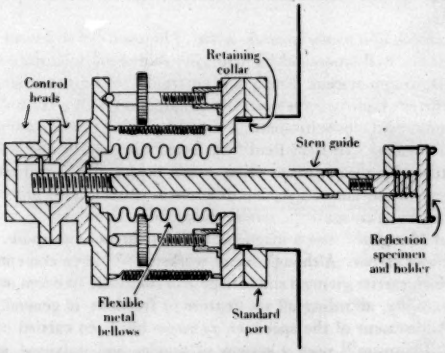


Figure 4. Specimen carrier.

diffraction camera is such that little is to be gained by the use of film except when several exposures are to be made without breaking vacuum, and even under these conditions a multi-plate camera as used by Trillat[27] is to be preferred.

Burgers[38] has described an instrument in which the electrons can be recorded by direct impact upon the photographic emulsion or alternatively by photographing the fluorescent screen with an ordinary lens camera outside the vacuum. Such indirect recording, however, leads to lack of definition and necessitates long exposures.

By using a Faraday cylinder Cochrane[44] has succeeded in recording the intensities as well as the dimensions of his diffraction patterns.

Vacuum technique. The efficient working of an electron diffraction camera is largely dependent upon the vacuum technique adopted. Rapid evacuation is desirable, whilst the introduction and removal of specimens and photographic plates necessitate vacuum-tight joints which are easily broken and remade; further-

more, in the case of the specimen carrier and camera shutter, joints which can be rotated without impairing the vacuum are essential.

Evacuation is usually effected by means of multi-staged mercury or oil diffusion pumps backed by rotary pumps, though Trillat[27], W. L. Bragg and Darbyshire[16] use Holweck molecular pumps. If a hot cathode is used, a single diffusion pump will serve cathode chamber and diffraction section, since both should be evacuated as highly as possible. In the case of the cold cathode tube, however, a pressure of the order of 10^{-2} mm. is required, necessitating differential pumping of the discharge tube and diffraction section. To obtain different degrees of vacuum in the two parts of the apparatus two diffusion pumps are generally used, but Finch and Quarrell[23], by employing an "injector" system, achieve the same result with a single pump and also eliminates the need for a liquid air trap.

Permanent vacuum-tight joints may be hard or soft-soldered, whilst semi-permanent joints are usually made with a low-vapour-pressure sealing compound such as picein. Large joints which are frequently made and broken are effected by means of rubber gaskets or plane ground metal surfaces lubricated with a suitable low-vapour-pressure grease[38]. Plane or conical ground surfaces are also used for rotating specimen carrier and shutter spindles.

The measurement of accelerating potential. The difficulties associated with the measurement of high voltages of the order of 30–80 kV. are well-known. The sphere-gap voltmeter affords a convenient method and is frequently used, but even under optimum conditions the error is not less than 2 per cent. The error in voltage measurement may be slightly reduced by the use of an electrostatic voltmeter or a microammeter in series with a high resistance, calibrated against the electron diffraction patterns given by a standard specimen.

Many workers have attempted to increase the accuracy of electron diffraction by the use of a comparison pattern in a manner analogous to the use of the iron arc reference in spectroscopy. Thus Finch and Quarrell[35] have described a double shutter camera (figure 5) such that the two halves of the stationary photographic plate can be exposed independently. Two specimens, one for examination and the other for reference, are mounted side by side on a specimen carrier of the type shown in figure 4, which is then so adjusted that when one specimen is in a suitable position in the electron beam, either simple rotation through 180° or a short predetermined traverse of the specimen carrier will bring the second specimen into a similar position. The first half of the plate is then exposed, whereupon the second specimen is brought into the beam and the exposure completed on the second half of the plate; the whole operation taking about 1 sec. All conditions affecting the production of the beam are unchanged throughout these operations and, owing to the strong bias on the focused beam, any adventitious change in the accelerating potential is betrayed by a displacement of one set of rings relative to the other.

Shishancow and Tartarinowa[40] have suggested using two convex specimens mounted face to face, the two patterns being recorded simultaneously. Owing to the nature of the pattern, however, the two sets of rings do not meet at a common boundary and are not described about a common centre.

Finch and Wilman[45] have successfully deposited their specimens upon gold and obtained composite patterns containing gold rings which were used for reference purposes.

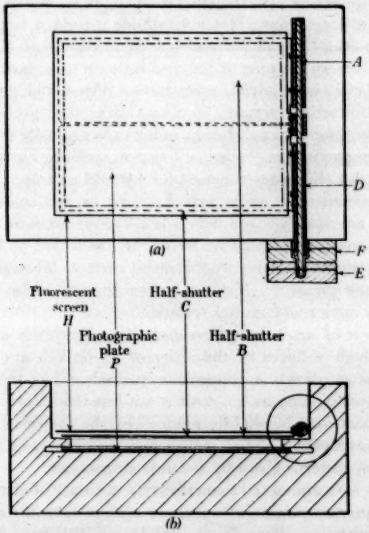


Figure 5. Details of double shutter camera (Finch and Quarrell, 1934).

Other workers have made use of a reference pattern by moving the photographic plate between the two exposures, but in this case the constancy of the accelerating potential cannot be proved.

REFERENCES

Part 2

(1) Morris, J. T. MacGregor and Mines, R. "Measurements in electrical engineering by means of cathode rays." *J. Instn elect. Engrs*, **63**, 1056 (1925).
(2) Dufour, A. *Onde élect.* **1**, 638 (1922).
(3) Wood, A. B. *J. Instn elect. Engrs*, **63**, 1046 (1925).
(4) Lee, E. S. *Gen. elect. Rev.* **31**, 404 (August 1928).
(5) Fortescue, Atherton and Cox. *J. Amer. Inst. elect. Engrs*, **48**, 449 (April 1929).
(6) Gábor, D. *Forschungsh. StudGes. Höchstspann Anlag.* Heft. 1 (Sept. 1927).
(7) Matthias, A., Knoll, M. and Knoblauch, H. "Kathodenstrahloszillographen liegender Banart." *Z. techn. Phys.* **11**, 276 (1930).
(8) Krug, W. *Elektrotech. Z.* **50**, 681 (May 9, 1929).

(9) NORINDER, H. "A cathode-ray oscillograph for electric line transients." *Tekniska Meddelanden Fran. Kungl. Vattenfallsstyrelsen*, Upsala (1927). Also *Tekn. Tidskr.* No. 31, *Electroteknik*, **8** (1925).
(10) ROGOWSKI, W. and FLEGLER, E. *Arch. Elektrotech.* **15**, 297 (1925).
ROGOWSKI, W., FLEGLER, E. and TAMM, R. *Arch. Elektrotech.* **18**, 479, 513 (1927).
(11) BERGER, K. *Bull. schweiz. elektrotech. Ver.* **19**, 292 (1928).
(12) BURCH, F. P. and WHELPTON, R. V. *J. Instn elect. Engrs*, **71**, 380 (1932).
(13) MILLER, J. L. *Electrician*, **109**, 805 (1932). See also Reference (29).
(14) ALLIBONE, T. E., HAWLEY, W. G. and PERRY, F. R. *J. Instn elect. Engrs*, **75**, 670 (Nov. 1934). See also discussion, *J. Instn elect. Engrs*, **76**, 236 (Feb. 1935).
(15) FINCH, G. I. and QUARRELL, A. G. *Proc. roy. Soc.* A, **141**, 398 (1933); *Proc. phys. Soc.* **46**, 148 (1934).
(16) PARKER SMITH, SZEGHÖ and BRADSHAW. *J. Instn elect. Engrs*, **76**, 656 (1935).
(17) BURCH, C. R. *Proc. roy. Soc.* A, **123**, 271 (1929).
(18) WHIPPLE, G. A. *Proc. phys. Soc.* **46**, 281 (1934).
(19) SIEKLE, VAN and BERKEY. *Trans. Amer. Inst. elect. Engrs*, **52**, 850 (1933).
(20) Personal communication to the author.
(21) BOEKELS, H. *Arch. Elektrotech.* **26**, 453 (1932).
(22) FREISEWINKEL, E. W. *Arch. Elektrotech.* **28**, 603 (1934).
(23) DODDS, J. M. *Arch. Electrotech.* **27**, 531 (1933).
(24) DICKS, H. *Arch. Elektrotech.* **25**, 523 (1931).
(25) Discussion. *J. Instn elect. Engrs*, **71**, 78 (June 1932).
(26) ROGOWSKI. *Elektrotech. Z.* **52**, 1245 (Oct. 1931).
(27) KNOLL, KNOBLAUCH and VON BORRIES. *Elektrotech. Z.* **51**, 966 (July 1930).
(28) BINDER, L. *Z. tech. Phys.* **9**, 379 (1930).
(29) MILLER, J. L. and ROBINSON, J. E. L. *J. Instn elect. Engrs*, **74**, 54 (1934).
(30) KNOLL. *Elektrotech. Z.* **53**, 1101 (Nov. 1932).
(31) BOEKELS and DICKS. *Arch. Elektrotech.* **27**, 134 (Feb. 1933).
(32) KLEMPERER and WOLFF. *Arch. Elektrotech.* **26**, 495 (1932).
(33) ACKERMANN. *Trans. Amer. Inst. elect. Engrs*, **49**, 467 (April 1930).
(34) GEORGE. *Trans. Amer. Inst. elect. Engrs*, **48**, 884 (1929).
(35) ROGOWSKI and BAUMGART. *Arch. Elektrotech.* **19**, 521 (1928).
(36) GRAUPNER. *Arch. Elecktrotech.* **28**, 477 (1934).
(37) GONDET and BEAUDOUIN. *Rev. gén. Élect.* **36**, 291 (Sept. 1, 1934).
(38) MALSCH and WESTERMANN. *Arch. Elektrotech.* **28**, 517 (Aug. 18, 1934).
(39) SOMMERFELD. *Arch. Elektrotech.* **20**, 607 (1928).
(40) ROGOWSKI, SOMMERFELD and WOLMAN. *Arch. Elektrotech.* **20**, 619 (1928).
(41) HARRIS. *Bull. U.S. Bur. Stand.* **12**, 87 (Jan. 1934).
(42) BUSH, H. *Arch. Electrotech.* **18**, 583 (1927).
(43) KNOLL. *Arch. Electrotech.* **28**, 128 (1934).
(44) ROGOWSKI. *Electrotech. Z.* **52**, 1245 (Oct. 1, 1931).
(45) KNOLL and v. BORRIES. *Z. tech. Phys.* **2**, 111, 491, 493 (1930).
(46) KNOLL. *Arch. Elecktrotech.* **26**, 132–4 (1932).
(47) BUSS, K. and PERNICK, A. *Arch. Elecktrotech.* **36**, 723–4 (Oct. 7, 1932).
(48) ROGOWSKI, W. *Elecktrotech. u. Maschinenb.* **51**, 249–54 (April 23, 1933).
(49) MCEACHRON, K. B. and GOODWIN, V. E. *Trans. Amer. Inst. elect. Engrs*, **48**, 953 (July 1929).
(50) FORTESCUE, ATHERTON and COX. *Trans. Amer. Inst. elect. Engrs*, **48**, 449 (April 1929).
(51) GROSS, I. W. and COX, I. A. *Trans. Amer. Inst. elect. Engrs*, **50**, 1118 (Sept. 1931).
(52) SPORN, P. and LLOYD, W. L. *Trans. Amer. Inst. elect. Engrs*, **50**, 1111 (Sept. 1931).
(53) NORINDER, H. *Trans. Amer. Inst. elect. Engrs*, **47**, 446 (April 1928).
(54) BOEKELS, H. *Arch. Elecktrotech.* **25**, 497 (1931).
(55) FIELDER, F. D. *Elect. J.* **28**, No. 3, p. 161 (March 1931).
(56) FOUST, C. M., KUEHNI, H. P. and ROHATS, N. *Gen. elect. Rev.* **35**, No. 7, p. 358 (July 1932).
(57) KRUG, W. *Elektrotech. u. Maschinenb.* **49**, No. 13, p. 233 (1931).

(58) KRUG, W. *Z. tech. Phys.* **13**, No. 8, p. 377 (1932).
(59) TAMM, R. *Arch. Elektrotech.* **19**, 235 (Jan. 25, 1928).
(60) GIROD, K. *Arch. Elektrotech.* **25**, 695 (1931).
(61) BAATZ, H., FREUNDLICH, M. and HOLZER, W. *Elektrotech. Z.* **53**, 696 (July 21, 1932).
(62) ECCLES, W. H. and JORDAN, F. W. *Radio Rev.* **1**, 145 (1919).
(63) GÁBOR, D. *Arch. Elektrotech.* **18**, 48 (1927).
(64) KNOLL, M. and FREUNDLICH, M. *Elektrotech. Z.* **53**, 669 (July 14, 1932).
(65) PEEK, H. *Arch. Elektrotech.* **25**, 452 (1931).
(66) ROGOWSKI, W. and WOLFF, O. *Arch. Elektrotech.* **21**, 645 (1929).
(67) NORINDER, H. *Tekniska Meddelanden Fran. Kungl. Vattenfallsstyrelsen* (June 15, 1931).
(68) DOWELL, J. C. and FOUST, C. M. *Trans. Amer. Inst. elect. Engrs*, **52**, 537 (June 1933).
(69) BELLASCHI, P. L. *Trans. Amer. Inst. elect. Engrs*, **52**, 544 (June 1933).
(70) Discussion on the above two papers.
(71) BURCH, F. P. *Phil. Mag.* Ser. 7, **13**, 760 (April 1932).
(72) GRÜNEWALD, H. *Electrotech. Z.* **55**, 505 (1934).
(73) FOUST, C. M. and GARDNER, G. F. *Gen. elect. Rev.* **35**, 324 (July 1934).
(74) BELLASCHI, P. L. *Trans. Amer. Inst. elect. Engrs*, **53**, 86 (1934).
(75) ROHATS, N. *Gen. elect. Rev.* **37**, 296 (June 1934).
(76) GRAUPNER, H. *Arch. Electrotech.* **26**, 725 (Oct. 7, 1932).
(77) WOOD, A. B. *J. Instn elect. Engrs*, **71**, 41 (1932).

PART 3

(1) BUSCH. *Ann. Phys.*, Lpz., **81**, 974 (1926); *Arch. Elektrotech.* **18**, 583 (1926).
(2) FARNSWORTH. *J. Franklin Inst.* **218**, 4, 411 (1934).
(3) JOHNSON. *J. opt. Soc. Amer.* **6** (1922).
(4) SCHERZER. *Z. Phys.* **80**, 3 and 4, 193 (1933).
(5) PICHT. *Ann. Phys.*, Lpz., **5**, 15, 926 (1932); *Z. tech. Phys.* **6**, 239 (1933).
(6) POSENER. *Z. Phys.* **80**, 11 and 12, 811 (1932).
(7) RUSKA. *Z. Phys.* **83**, 9 and 10, 684 (1933).
(8) OLLENDORF and WENDT. *Z. Phys.* **76**, 9 and 10, 655 (1932).
(9) GLASER. *Z. Phys.* **80**, 7 and 8, 451 (1933); **81**, 9 and 10, 647 (1933); **83**, 1 and 2, 104 (1933); *Ann. Phys.*, Lpz., **5**, 18, 557 (1933).
(10) KNOLL and RUSKA. *Ann. Phys.*, Lpz., **12**, 5, 607 (1932); **12**, 6, 641 (1932).
(11) BRÜCHE. *Naturwissenschaften*, **20**, 3, 49 (1932).
(12) BRÜCHE and JOHANNSON. *Naturwissenschaften*, **20**, 21, 353 (1932).
(13) JOHANNSON and SCHERZER. *Z. Phys.* **80**, 3 and 4, 183 (1932).
(14) HAMACHER. *Arch. Elektrotech.* **26**, 215 (1932).
(15) GRAUPNER. *Arch. Elektrotech.* **26**, 725 (1932).
(16) BRÜCHE. *Z. Phys.* **78**, 1 and 2, 26 (1932); *Arch. Elektrotech.* **27**, 266 (1933).
(17) ZWORYKIN. *J. Franklin Inst.* **215**, 535 (1933).
(18) RUSKA and BORRIES. *Z. Phys.* **76**, 9 and 10, 649 (1932).
(19) JOHANNSON. *Ann. Phys.*, Lpz., **18**, 4, 385 (1933); **21**, 3, 274 (1934).
(20) KNOLL and RUSKA. *Z. tech. Phys.* **12**, 389 (1931); *Z. Phys.* **78**, 5 and 6, 318 (1932).
(21) KNOLL, HOUTERMANS and SCHULZE. *Z. Phys.* **78**, 5 and 6, 340 (1932).
(22) RUSKA. *Z. Phys.* **87**, 9 and 10, 580 (1934).
(23) JOHANNSON and KNECHT. *Z. Phys.* **86**, 5 and 6, 367 (1933).
(24) KNECHT. *Ann. Phys.*, Lpz., **5**, 20, 161 (1934).
(25) v. ARDENNE. *Z. Phys.* **86**, 11 and 12, 802 (1933).
(26) BRÜCHE and JOHANNSON. *Phys. Z.* **33**, 898 (1932).
(27) RICHTER. *Z. Phys.* **86**, 11 and 12, 697 (1933).
(28) BRÜCHE and JOHANNSON. *Ann. Phys.*, Lpz., **15**, 2, 145 (1932); *Z. Phys.* **84**, 1 and 2, 56 (1933).
(29) KNOLL and LUBSZYNSKI. *Phys. Z.* **34**, 671 (1933).

(30) BRÜCHE. *Z. Phys.* **86**, 7 and 8, 448 (1933).
(31) RUSKA. *Z. Phys.* **83**, 7 and 8, 492 (1933).
(32) RUSKA and BORRIES. *Z. Phys.* **83**, 3 and 4, 187 (1933).
(33) BRÜCHE and KNECHT. *Z. tech. Phys.* **15**, 461 (1934).
(34) MARTON. *Nature*, Lond., **133**, 911 (1934); *Bull. Acad. Belg. Cl. Sci.* **20**, 439 (1934); *Phys. Rev.* **46**, 527 (1934).
(35) RUSKA. *Z. Phys.* **89**, 90 (1934).
(36) BRÜCHE and SCHERZER. *Geometrische Elektronenoptik* (1934). Berlin: Springer.
(37) BRÜCHE. *Ann. Phys.*, Lpz., **16**, 4, 37 (1933).
(38) —— *Z. tech. Phys.* **2**, 49 (1933).
(39) BRÜCHE and SCHERZER. *Z. tech. Phys.* **14**, 11, 464 (1933).
(40) v. ARDENNE. *Z. Phys.* **88**, 3 and 4, 251 (1934).
(41) HAMACHER. *Arch. Elektrotech.* **27**, 47 (1933); **27**, 121 (1933).
(42) WENDT. *Ann. Phys.*, Lpz., **5**, 17, 445 (1933).
(43) FRANK. *Z. Phys.* **80**, 4 (1933).
(44) VALLE. *Elettrotecnica*, **20**, 697 (1933).
(45) RICHTER. *Phys. Z.* **34**, 457 (1933).
(46) SCHERZER. *Z. Phys.* **82**, 11 and 12, 697 (1933).
(47) BOBSKOVSKY and FRENKEL. *Phys. Z. Sowjet.* **5**, 3, 464 (1934).
(48) ENGEL. *Z. Phys.* **79**, 3 and 4, 231 (1933).
(49) ROGOWSKI and GRAUPNER. *Arch. Elektrotech.* **26**, 807 (1932).
(50) MARTIN. *J. Televis. Soc.* **1**, 2, 377 (Dec. 1934).
(51) BENHAM. *J. Instn elect. Engrs*, **75**, 388 (1934).
(52) BARBER. *Proc. Leeds phil. lit. Soc.* **2**, 427 (1933).
(53) DAVISSON and CALBICK. *Phys. Rev.* **38**, 585 (1931); **42**, 580 (1932).
(54) WATSON. *Phil. Mag.* **3**, 849 (1927).
(55) BENNETT. *Phys. Rev.* **46**, (June 15, 1934).

PART 4

(1) THOMSON, G. P. *Proc. roy. Soc.* A, **117**, 600 (1928); *Nature*, Lond., **119**, 820 (1927).
(2) REID, A. *Proc. roy. Soc.* A, **119**, 664 (1928).
(3) THOMSON, G. P. and FRASER, C. G. *Proc. roy. Soc.* A, **128**, 641 (1930).
(4) KIKUCHI, S. *Jap. J. Phys.* **5**, 83 (1928).
(5) MUTO, T. and YAMAGUTI, T. *Proc. imp. Acad. Japan*, **5**, 122 (1929).
(6) MATUKAWA, K. and SHINOHARA, K. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, No. 424, p. 21 (1933).
(7) WIERL, R. *Ann. Phys.*, Lpz., **8**, 521 (1931).
(8) PONTE, M. *Ann. Phys.*, Paris, **13**, 395 (1930).
(9) KIRCHNER, F. *Phys. Z.* **31**, 772 (1930).
(10) SCHOBITZ, E. *Phys. Z.* **32**, 37 (1931).
(11) RUPP, E. *Ann. Phys.*, Lpz., **10**, 927 (1931).
(12) EISENHUT, O. and KAUPP, E. *Z. Elektrochem.* **37**, 466 (1931).
(13) TRILLAT, J. J. and HIRSCH, T. v. *C.R. Acad. Sci.*, Paris, **193**, 649 (1931).
(14) KIRCHNER, F. *Ann. Phys.*, Lpz., **11**, 741 (1931).
(15) MEIBOM, R. v. and RUPP, E. *Ann. Phys.* Lpz., **13**, 725 (1932).
(16) BRAGG, W. L. and DARBYSHIRE, J. A. *Trans. Faraday Soc.* **28**, 522 (1932).
(17) THIBAUD, J., TRILLAT, J. J. and HIRSCH, T. v. *J. Phys. Radium*, **3**, 314 (1932).
(18) SHINOHARA, K. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, No. 368–9, 223 (1932).
(19) MONGAN, C. *Helv. phys. Acta*, **5**, 341 (1932).
(20) BRAUNE, H. and KNOKE, S. *Z. phys. Chem.* B, **21**, 297 (1933).
(21) KIKUCHI, S. and NAKAGAWA, S. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, No. 430, 80 (1933).
(22) MEIBOM, R. v. and RUPP, E. *Zeits. f. Physik.* **82**, 690 (1933).
(23) FINCH, G. I. and QUARRELL, A. G. *Proc. roy. Soc.* A, **141**, 398 (1933).

(24) Hendricks, S. B., Maxwell, L. R., Mosley, V. L. and Jefferson, M. E. *J. chem. Phys.* **1**, 549 (1933).
(25) Maxwell, L. R. *Phys. Rev.* **44**, 73 (1933).
(26) Dornte, R. W. *J. chem. Phys.* **1**, 586 (1933).
(27) Trillat, J. J. *Trans. Faraday Soc.* **29**, 995 (1933).
(28) Darbyshire, J. A. *Phil. Mag.* **16**, 761 (1933).
(29) Richards, D. A. *Phil. Mag.* **16**, 778 (1933).
(30) Kosman, M. and Alichanian, A. *Phys. Z. Sowjet.* **4**, 551 (1933).
(31) Laschkarew, W. E. and Usyskin, I. D. *Z. Phys.* **85**, 618 (1933).
(32) Nelson, H. R. *Phys. Rev.* **44**, 717 (1933).
(33) Fricke, R. and Luke, J. *Z. phys. Chem.* B, **23**, 330 (1933).
(34) Emslie, A. G. *Phys. Rev.* **45**, 43 (1934).
(35) Finch, G. I. and Quarrell, A. G. *Proc. phys. Soc.* **46**, 148 (1934).
(36) Lassen, H. *Phys. Z.* **35**, 172 (1934).
(37) Yamaguti, T. *Proc. phys.-math. Soc. Japan*, **16**, 95 (1934).
(38) Burgers, W. G. and Basart, J. C. M. *Physica*, 's Grav., **1**, 543 (1934).
(39) Thomson, G. P. *Phil. Mag.* **17**, 1058 (1934).
(40) Shishancow and Tartarinowa. *Nature*, Lond., **133**, 686 (1934).
(41) Myers, F. E., Byrne, J. F. and Cox, R. T. *Phys. Rev.* **46**, 777 (1934).
(42) Darbyshire, J. A. and Cooper, E. R. *Trans. Faraday Soc.* **30**, 1038 (1934).
(43) Trillat, J. J. *Kolloidszchr.* **69**, 378 (1934).
(44) Cochrane, W. *Phil. Mag.* **18**, 956 (1934).
(45) Finch, G. I. and Wilman, H. *J. chem. Soc.* p. 751 (1934).

X RAYS

By THE X-RAY STAFF, PHYSICS DEPARTMENT, NATIONAL PHYSICAL LABORATORY

Edited by G. W. C. Kaye, O.B.E., M.A., D.Sc.

§ 1. PRODUCTION OF X RAYS

By G. E. Bell, M.Sc.

High-tension generators. The question of the most suitable type of high-voltage generator for X-ray purposes has been of practical significance for many years for industrial research as well as for medical purposes. Modern generators usually consist of a high-tension transformer the output voltage of which is rectified by means of hot cathode valves, and often smoothed by means of suitable condensers. When very high voltages are required the output voltage of the transformer is frequently increased by means of suitably disposed condensers and rectifiers. Recent developments in both high-power and high-voltage X-ray tubes have called attention to certain defects in condenser equipment which may limit both the voltage attainable and the power output available. In particular it has been known for some time that the output voltage of generators of the condenser type falls considerably with the load, and that rectifying valves are often required to pass much higher currents than those delivered to the X-ray tube. An oscillographic study of high-tension generators of the types commonly used for X-ray purposes made by Kaye and Bell[1] has revealed the causes of some of the anomalies in the behaviour of such circuits and showed the advantages of using equipment having full-wave rectification. The properties of constant voltage generators of the condenser type have been investigated both theoretically and practically by Bell[2] who showed that the electrical properties of such circuits are largely influenced by the resistance of the secondary circuit, which is responsible for the fall in voltage, but at the same time serves to limit the peak value of the current passed by the rectifying valves.

The variation of X-ray output with tube voltage, and the influence of the wave form of the exciting voltage on the quality and quantity of X rays emitted by a particular tube has also been the subject of several investigations. It seems certain from the work of Kulenkampff and others that, under conditions in which full allowance is made for absorption and other disturbing factors, the X-ray output (as measured by an absolute method) of a particular tube at a fixed current is proportional to the square of the peak voltage of excitation. Usually, however, the "intensity" of X rays is not measured absolutely but is specified by the degree of ionization produced in an ionization chamber of suitable design. This ionization

results from that portion of the beam which is absorbed in the chamber so that, unless corrections are applied, such measurements determine, not the energy or intensity of the beams, but a quantity resting partly on the intensity and partly on the quality. This quantity, which is often directly associated with the "dose" in therapeutic measurements is conveniently termed the "dosage-rate" of irradiation. Kaye and Bell[3], working over the voltage range 50 kV. to 150 kV. showed that, for hot-cathode X-ray tubes of normal design, the dosage rate (D) could be related to the voltage (V) through the equation $D=kV^n$, and that n varies from nearly unity for X rays which are practically unfiltered to 4 or 5 for radiation filtered by 1 mm. of copper. Thoreus[4] also found n to be nearly unity for lightly filtered radiation. Roberts[5] investigated the corresponding problem for high voltage radiations and over the voltage range 100 to 360 kV. peak found that no single value of n was valid over the entire voltage range. For lightly filtered radiation he found values of n less than 2.

As regards the influence of the wave-form of the exciting voltage, Behnken, Hull, Glocker and Nasledow and Kacura showed that under like conditions of peak voltage and tube current, an X-ray tube excited by a constant voltage gives about 50 per cent greater dosage rate than when excited by a pulsating voltage. Niemann and Zeyss[6] using radiation filtered by 1 mm. of aluminium found that constant voltage excitation gave about double the output of pulsating voltage excitation. Kaye and Bell[3] found that, for unfiltered radiation the ratio is about 1·8, and that as the filtration is increased the superiority of the constant voltage generator becomes more marked. L. S. Taylor[7] and his colleagues obtained similar results and stated that the X-ray dosage-rates were more closely related to r.m.s. voltages than to peak voltages. The various investigations show that X rays excited at constant voltage are more penetrating than those excited by pulsating voltages of the same peak value.

High-power X-ray Tubes. The most serious difficulty in the design of high-power X-ray tubes is that of dissipating the heat generated by the impact of the cathode rays on the target. The efficiency of an X-ray tube is very low—the latest determination by Gaertner[8] giving an efficiency of X-ray production of only about 0·3 per cent—so that in order to obtain very intense beams of X rays, whether for physical or medical purposes, tubes must be loaded to 5 or 10 kw., and this power is incident on a focus of only a few square millimetres in area. From the point of view of design, X-ray tubes may be divided fairly sharply into two types, namely those designed for short time exposures with long intervals for cooling and those designed for continuous operation, the latter being the more important from the point of view of the physicist. In tubes of the first type the target is usually very massive and so designed as to have a high thermal capacity, while tubes of the second type usually rely on cooling by the forced circulation of water or oil.

Apart from the problem of dissipating the total heat generated, there is the problem of providing a sufficiently fine focus within the limits set by the melting point and thermal conductivity of the target material. In practice it is the effective size of the focal spot and not its real size which is important, a fact turned to

account by Goetz in his "line focus". In line-focus tubes the filament takes the form of a straight helix mounted within an aperture which permits a roughly rectangular beam of electrons to fall on the target. The rectangular focal spot is viewed end-on at such an angle that the effective focal spot is roughly square. In this way an effective focus say 3 mm. square is obtained from an actual focus about 9 by 3 mm., and the permissible loading can be increased about three times.

Calculations on the limits of the power input to X-ray tubes have been made by Bouwers[9] who specially considered short-time loadings and by Müller[10] who considered continuous loadings. Bouwers studied theoretically and practically the rise in temperature with time of the anode of a line focus tube, the target of which consisted of a tungsten sheet on a copper base, and showed that, for very short exposures (0·04 sec.), the temperature rises proportionally to the load and to the square root of the loading time, while for longer exposures the temperature increases less rapidly than the square root of the loading time.

Müller considered the steady state of a circular focus the heat inflow to which is given by

$$\text{Inflow} = Ae^{-Br^2}$$

an equation expressing the fact that the inflow of heat reaches a maximum at the focus centre ($r=0$) and drops rapidly for increasing distances from the focus centre. Defining the radius of the focus as the distance at which the inflow drops to half its maximum value, Müller finds that the maximum permissible loading W_{max} of a stationary anode of material having a melting point $T°$ C. and a thermal conductivity k (gm.-cal. $sec.^{-1}$ $cm.^{-1}$) is given in watts by the equation $W_{max} = 17{\cdot}8\,(T - T_0)\,k$. For a focus of diameter 1 mm. this gives the following approximate values of W_{max} in watts:

Copper target	760
Iron target	160
Tungsten target	780
Gold target	500

The principal uncertainty in such calculations of the input limits lies in the value of k, which is a function of the temperature.

Another way of increasing the permissible loading of an X-ray tube is to keep the target moving and so to distribute the heat. The theoretical aspects of this problem have been studied by Bouwers[11] and by Müller[12], the former from the point of view of instantaneous loading and the latter from the point of view of continuous loading.

Bouwers found that, for a given size of linear focal spot and for very short exposures the maximum permissible power input to a rotating target is greater than that to a stationary target in the ratio

$$\frac{W_{max}\ (\text{rotating})}{W_{max}\ (\text{stationary})} = \sqrt{\frac{2\pi nrt}{f}},$$

where n is the number of revolutions per second, r the radius of the focus path, f

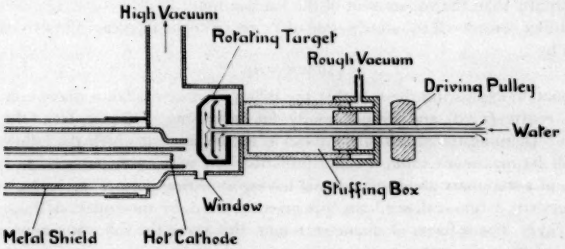
the breadth of the focal spot and t the exposure time, while Müller found, for the continuous loading of a circular focal spot the ratio

$$\frac{W_{\max}\text{ (rotating)}}{W_{\max}\text{ (stationary)}} \fallingdotseq \sqrt{\frac{2\pi nrf \cdot \rho c}{k}},$$

where ρ is the density of the target material, c its specific heat and k its thermal conductivity. From Müller's formula it may be shown that if a copper target of radius 7·5 cm. rotates at about 2500 rev. per min., the permissible energy input with a rotating target is increased tenfold.

These formulae show that the power ratio is proportional to the square root of the speed of rotation and of the radius of the focus path, and both these conclusions have been verified experimentally.

High-power X-ray tubes employing rotating anodes have been described by Bouwers[11] and by Clay[13]. Bouwers uses a sealed-off self-protected tube in which



Müller rotating-target X-ray tube

Figure 1. Laboratory X-ray tube with rotating anode (Müller and Clay).

the target consists of a tungsten-faced cylindrical copper block mounted so as to be capable of rotation. Outside the tube is mounted the stator of a three-phase motor. The rotating field of the stator produces eddy currents in the copper anode which cause the anode to rotate, stator and anode constituting an induction motor. Tubes of this type operate most satisfactorily on three-phase high tension supplies with full-wave rectification. With a focal area of about 15 mm^2 they are capable of carrying a load of 30 kw. for 1/20 sec.—this being about 8 times that permissible with a stationary focus of the same size and shape.

The tube described by Clay, which is designed primarily for physical research, and is of the continuously-evacuated water cooled type, capable of operation continuously with a tube current of 150 mA. and a tube potential of 30 kV. (peak). The theoretical maximum continuous loading of the tube, which has a copper target, is 40 kW. according to Müller's formula, but practical limitations such as cracking and roughening of the target and imperfections of cooling reduce the loading to about 5 kW. A diagrammatic section of the tube is given in figure 1. The target consists of a hollow copper disc about 10 cm. in diameter mounted at

one end of a hollow shaft which runs in ball bearings, passes out through a stuffing-box and is driven by a belt and pulley, while water circulates through the shaft and the inside of the disc. The enclosure in which the disc runs is exhausted by a fast diffusion pump to a vacuum suitable for the operation of a hot filament as a source of cathode rays.

Perhaps the most interesting mechanical feature of the tube is the device adopted to prevent appreciable quantities of air leaking into the discharge chamber along the driving shaft. The stuffing-box already mentioned, which contains suitable gaskets of lenticular cross-section pressed into fairly close contact, is connected to a rough vacuum by the tube shown. The use of the intermediate vacuum in the stuffing-box prevents any appreciable leakage of air from the atmosphere along the driving shaft and into the discharge chamber.

§ 2. IONIZATION IN GASES, ABSORPTION AND SCATTERING OF X RAYS

By I. Backhurst, M.Sc.

Ionization in gases. A method of obtaining a beam of positive ions up to 10 mA. in intensity has been developed by Fowler and Gibson[1]. The source is an arc, working on an alternating supply of 200 to 400 V. and maintained between a heated tungsten wire cathode and a double copper anode in a gas at a pressure below atmospheric. Gas is continuously admitted to the arc chamber near the anode, as it is constantly being lost through the tubular aperture, or "extractor", connecting the cathode end of the arc chamber to the ion chamber. This occurs since the latter is maintained at a sufficiently low pressure to enable an accelerating potential of about 20 kV. to be applied to the ions in transit between the two chambers via the extractor. Magnetic resolution of the beam from a hydrogen arc showed the presence of H_1^+, H_2^+ and H_3^+ ions, the beam at 20 mm. of mercury pressure consisting almost entirely of H_1^+ ions, while at 72 mm. of mercury it contained only H_1^+ ions.

Hydrogen ions may now be accelerated up to energies corresponding to 5 million electron volts, as a consequence of improvements made by Lawrence and Livingston[2] in their original apparatus. Two semicircular hollow electrodes are situated in a chamber containing hydrogen at a pressure of 10^{-4} to 10^{-5} mm. Hg and electronic emission from a heated filament ionizes the gas near the centre of the diametral region between the electrodes. A uniform magnetic field is arranged normal to the plane of the electrodes and a high frequency oscillating potential of 20 kV. is applied between them. The ions are made to spiral around in synchronism with the field, gaining increments in energy corresponding to the applied voltage each time they cross the diametral region, and finally emerging at the periphery of an electrode as a definite beam, homogeneous to within about 5 per cent and of about 1/3 microamp. intensity.

The production of beams of uncharged atoms of low velocity has been studied by Beeck[3], who has developed a method in which the first step is the production, by means of electrons from a heated filament, of ions in a gas at a pressure of about

10^{-3} mm. Hg inside a tube of 1·5 mm. diameter. The electrons are allowed to acquire a velocity less than that corresponding to the second ionization potential of the gas and slow, singly ionized, positive ions, together with slow neutral particles and some electrons, emerge from the tube into a larger chamber, where an electric field accelerates the ions and stops the electrons. Two lateral fields of opposite sign are then employed in succession to displace the ion beam sideways, thereby enabling the latter to be separated from the neutral particles by means of a slit system. The homogeneous ion beam so obtained is allowed to pass into a chamber, in which the gas pressure is higher, and those ions, that become neutralized without suffering much deflection, pass through a further slit system, where an electric field is used to cut off any ions that have escaped neutralization. Using a beam of argon atoms obtained in this way, Beeck and Wayland[4] have found that the ionization, per atom of beam per unit path length, produced in argon, is about four times that caused by a beam of positive potassium ions in argon and therefore still further exceeds that produced by a beam of positive argon ions in argon.

Absolute measurements of the ionization energy ϵ required to produce an ion pair in argon, when the primary ionizing agent is an X-ray beam from a copper anticathode, have been made by Gaertner[5], who finds $\epsilon_{argon} = 28{\cdot}4 \pm 0{\cdot}5$ V. Using known ratios of ϵ_{air} to ϵ_{argon}, the value deduced for ϵ_{air} is considered by Gaertner to differ from the accepted value of 32·2, found when cathode rays are the cause of ionization, by an amount greater than experimental error. The explanation suggested is that an atom may be able to absorb a radiation quantum by a direct process of conversion of the latter into kinetic energy of the atom as a whole, the emission of neither electronic nor X radiation being involved.

Absorption of X rays. When the usual crystal reflection method is used to produce a homogeneous X-ray beam of short wave-length, it is necessary, on account of the smallness of the angle of reflection, to use very narrow spectrometer slits and the X-ray intensity may then be inconveniently low for absorption measurements. Hahn[6] has investigated the possibility of overcoming this difficulty by using the double crystal spectrometer, with which, under normal conditions of adjustment, the degree of homogeneity of the twice reflected beam depends mainly on the state of perfection of the crystals and not at all on slit width, although it does depend to some extent on slit height. The $K\alpha_1$ line of tungsten ($\lambda = 0{\cdot}2086$ A.) was selected and X-ray intensities, with and without the interposition of the absorbing material, compared by means of ionization produced in a chamber containing air at 30 atmospheres pressure. The materials examined were paraffin, graphite, aluminium, copper, silver, tantalum, tungsten and lead. Using only one crystal, as in the usual way, various wave-lengths in the range 0·140 to 0·230 A. were also employed. The two methods were found to give consistent results and confirm the absorption coefficient determinations of previous observers.

The part of the absorption coefficient due to scattering is important for elements of low atomic number, at these wave-lengths, and Hahn has estimated this by making the assumption that, at a fixed wave-length, the logarithm of the part of the electronic absorption coefficient μ_a/Z, not due to scattering, is a linear function

of the logarithm of the atomic number Z. It is also assumed that σ_s, the part of μ_a/Z due to scattering, may be regarded as nearly enough independent of Z, so that by trial it is possible to find a value of σ_s to make the graph of $\log([\mu_a/Z]-\sigma_s)$ against $\log Z$ a straight line. Values of σ_s so found at 0·2086, 0·1700 and 0·1400 A. agreed closely with values calculated from the Klein-Nishina theory, which differ, in this region of wave-length, about 2 per cent from those deduced from the Compton or Breit-Dirac theories. It is interesting to note that the absorption in hydrogen, as deduced from the results for graphite and paraffin, appeared to be wholly due to scattering.

An entirely different solution of the problem of obtaining a homogeneous beam has been utilized by Grosskurth[7], who has followed the lines of a method, described in some detail by Kustner[8], in which the homogeneous beam used is not actually separated from a band of "white" radiation. Some of the secondary radiation proceeding from an element R, placed in the path of a beam of primary rays from an X-ray tube T, is received by an ionization chamber I. This secondary radiation, with a suitable X-ray tube voltage, will consist of characteristic K radiation as well as of scattered radiation. A sheet of copper S, sufficiently thick to reduce the K radiation to about 1 per cent of its initial intensity is placed, first, between R and I, then between R and T. If conditions are such that the scattered part of the secondary radiation is nearly enough of the same average "hardness" as the primary, then its effect at I will be the same for both positions of S. The intensity of the K radiation at I will, on the contrary, be different in the two cases, so that the measured difference in total intensity at I may be regarded as the intensity of a beam of K radiation from R. The degree of homogeneity of the latter may be improved by using, as a filter between R and I, an element having its K absorption discontinuity at a wave-length intermediate between the α and β components of the K radiation from R.

Grosskurth's results confirm those of others with regard to the impossibility of obtaining an accurate universal formula of the type $\mu/\rho = a \cdot Z^b \cdot \lambda^c$, where μ/ρ is mass absorption coefficient, Z is atomic number and λ is wave-length. The magnitude of the K absorption jump, for elements of high atomic number, was found to be substantially lower than that deduced from the atomic energy levels by means of Jönsson's equation[9].

Employing the more usual methods, absorption measurements have also been made recently by White[10] and Carr[11]. White has studied xenon, with λ between 0·18 and 1·5 A., while Carr has examined the L discontinuities of bismuth. Both are able to represent their results by formulae of the type $\mu/\rho = a \cdot \lambda^b + c$, where a, b and c remain constants so long as λ does not cross a K or L discontinuity.

Scattering of X rays. As mentioned in the 1934 *Report*, the phenomena of scattering are usually complicated by interference between radiation scattered from different atoms of the scattering substance. The theoretical expressions for pure atomic scattering may in some cases be modified so as to represent fairly accurately the actual scattering observed when such interference takes place. Measurements made by Williams[12] of the "diffuse" scattering from a single

crystal of sodium fluoride have been used by Harvey, Williams and Jauncey[13] to test the accuracy of two theoretical expressions, based on wave-mechanics, that have been developed, one by Wentzel[14] and the other by Waller and Hartree[15].

Williams employed as incident radiation a monochromatic X-ray beam obtained by selecting MoKα rays from a primary beam with the aid of a rock salt crystal. Great care was taken to avoid the entrance into the ionization chamber of any regularly reflected rays from the sodium fluoride. The manner in which the relation between the diffusely scattered intensity so found and the angle of scattering is compared with theory may be briefly indicated as follows. On both the above theories the coherent scattering per electron, in terms of the Thomson value per electron, is given by $s_1 = f^2/Z$, where $f = \sum_{j=1}^{z} E_{jj}$, and $E_{jk} = \int \psi_j \psi_k^* e^{i\sigma\mu r} . r^2 . dr . d\mu . d\phi$, where $\mu = \cos\theta$, $\sigma = 4\pi \left(\sin\frac{\phi}{2}\right) \Big/ \lambda$, ϕ is the angle of scattering and ψ_j is the wave function associated with the jth electron, Z being the atomic number.

Wentzel's formula for the incoherent scattering is $s_2 = 1 - \{(\sum_{j=1}^{z} E_{jj}^2)/Z\}$, while that of Waller and Hartree is $s_2 = 1 - \{(\sum_{j=1}^{z} E_{jj}^2)/Z\} - \{(\sum_{j=1}^{z} \sum_{j+k} |E_{jk}|^2)/Z\}$. The extra term in the latter expression arises from taking into account the Pauli exclusion principle, which restricts the number of possible electron transitions between different energy states and so reduces the calculated radiation intensity. The total scattered intensity per electron, in terms of the Thomson scattering, is then $s_1 + s_2 . (1 + \alpha . \text{vers}\,\phi)^{-3}$ on bringing in the relativity correction (see *Report* 1934).

For a crystal of the two kinds of atoms 1 and 2, Harvey, Williams and Jauncey, on the basis of Wentzel's theory, use for $(s_1 + s_2)$ the following expression:

$$1 - \frac{\Sigma (E_{jj}^2)_1 + \Sigma (E_{jj}^2)_2}{Z_1 + Z_2} + \frac{f_1^2 + f_2^2 - F_1^2 - F_2^2}{Z_1 + Z_2},$$

and on the basis of the theory of Waller and Hartree (these authors write f_{jk} instead of E_{jk}) a quantity less than this by the amount

$$\frac{\Sigma |E_{jk}|_1^2 + \Sigma |E_{jk}|_2^2}{Z_1 + Z_2},$$

F_1 and F_2 are the atomic structure factors taken to include the temperature effect and represent the effect of interference between radiation from different atoms of the crystal, i.e. F_1 and F_2 allow for the intensity diverted to the regularly reflected spectra. With sodium fluoride a simplification occurs, in that the Na^+ and F^- ions are both neon like, so that in the calculations the wave functions for neon may be used. It is found that Williams' experimental results are well represented when the extra term due to the Pauli exclusion principle is taken into account, but otherwise the calculated value of the total scattering is too high by an amount greater than the estimated maximum experimental error.

A somewhat similar test of recent theory has been subsequently made by Jauncey and Claus[16], who have repeated Harvey's measurements of the diffuse

scattering from sylvine. Harvey's results gave a maximum of intensity for (s_1+s_2) of 2·95 at $\left(\sin\frac{\phi}{2}\right)\Big/\lambda = 0{\cdot}365$ and did not give close agreement with theory, as, using James and Brindley's "f" values, the maximum value calculated for (s_1+s_2) is 2·46. Jauncey and Claus measured the scattering for angles between 10° and 90° for an average wave-length, as determined by half-value absorption, of 0·38 A. The maximum value of (s_1+s_2) was found to be 2·31 at 25°, for which $\left(\sin\frac{\phi}{2}\right)\Big/\lambda$ is 0·57, thus showing satisfactory agreement with the corresponding theoretical value for (s_1+s_2) of 2·36.

Harvey[17] has also studied the effect of interference between radiation scattered from neighbouring atoms by measuring the intensity of scattering from nitrogen for angles less than 8°, at pressures of 60, 80 and 100 atmospheres. Rays from a molybdenum anode were employed, after filtering with a zirconium oxide screen, and a krypton ionization chamber was used in conjunction with a Compton electrometer. The experimental difficulties were considerable, the background radiation being sometimes as much as 80 per cent of the total measured effect. It was possible to measure only the relative intensity at different angles, so that in making comparison with a theory of Debye's[18], the ordinates of the experimental curves showing angular intensity distribution were adjusted to give the best fit with the corresponding theoretical curves at peaks of maximum intensity. The experimental curve at each pressure showed a peak at almost the identical position required by theory, but the slope of the curve on the small angle side of the peak was in each case steeper than the theoretical.

§ 3. X-RAY (AND GAMMA-RAY) QUANTITY OR DOSE. PROTECTION

By W. Binks, M.Sc.

Quantity or dose. During the years preceding and immediately following the adoption of the röntgen in 1928 as the international unit of X-ray quantity or dose, X-ray measurements were confined to a limited range of wave-lengths, corresponding to exciting voltages of about 30 to 200 kV. With the advent of low-voltage (grenz-ray) and high-voltage X-ray tubes, there followed the construction of ionization chambers suitable for X-ray measurements in röntgens over a much wider range of wave-lengths. Such chambers belonged either to the free-air type or to the air-wall type.

The difficulty in the measurement of grenz radiation (that is, of soft X rays produced at voltages up to about 12 kV.) lies in the high absorption of the radiation, with a consequent alteration of the quality. In this region, air-wall chambers are unsuitable since the absorption correction even in the case of the thinnest walls is fairly large and cannot be determined accurately. Jaeger[1] and also Taylor and Stoneburner[2] have accordingly used small free-air chambers. The latter authors reduced the size of the chamber by means of a guarded-field arrangement, thereby

considerably reducing the absorption in the air between the X-ray source and the ionized air volume within the chamber.

Another difficulty revealed itself when the measurements were extended to high-voltage X rays. The definition of the röntgen includes the condition that full utilization shall be made of the energy of the secondary electrons generated in a definite volume of air. Owing to the larger ranges of the electrons liberated by hard X rays, the electrons proceed outside the volume in which they are generated and consequently the major portion of their energy is not measured. It is however possible to compensate this loss of energy. This compensation can be effected (*a*) in the case of free-air chambers, by making the air path between the X-ray tube and the measured air volume within the chamber at least equal to the maximum forward electron range in air and (*b*) in the case of air-wall chambers, by having the walls of a thickness at least equal to the forward electron range in the material in question.

Measurements of high-voltage X rays have been made by Mayneord and Roberts[3] and Jaeger[4] by comparing the readings of large free-air and thick air-wall chambers. The former authors used a rectangular free-air chamber having parallel-plate electrodes 20 cm. apart, the aperture at the front of the chamber being about 35 cm. from the centre of the ionized volume. Their air-wall chamber had outer walls of graphite 4 mm. thick. They found that a deviation between the chambers commenced with X radiation excited at 260 kV. and filtered through 4·4 mm. of copper. This deviation was attributed to the fact that the electron ranges at voltages in excess of about 250 kV. are greater than 35 cm., which is the distance from the chamber aperture to the centre of the collecting electrode; in other words that the free-air chamber used was too small.

Jaeger worked with a similar size of free-air chamber, this being cylindrical and having a diameter of 25 cm. and length of 50 cm. This was found to agree with a thick air-wall chamber for all qualities of X radiation up to the hardest employed (500 kV., and 3 mm. copper filter). The quality of this radiation may alternatively be expressed as having a half-value layer of 6·0 mm. copper. At this quality Jaeger finds that the free-air chamber gives a 2 per cent higher reading than the air-wall chamber, whereas Mayneord and Roberts find that it is 6 per cent lower and moreover that the readings of the two chambers diverge still further as the half-value layer of the radiation is increased. This apparent contradiction of results remains to be reconciled.

In the region of gamma radiation from radium, which may be considered as equivalent to X radiation generated at voltages between 250 and 2200 kV., the difficulties of measurement are much greater, not only because of the large range of the electrons (about 4 metres in air at atmospheric pressure) but also because of the importance of scattered radiation. As a result, the measurement of gamma-ray dosage is lagging behind that of X-ray dosage. At first, the chosen gamma-ray units, such as the milligramme-element-hour and the millicuries destroyed, were defined in terms of the source, since it appeared that the emitted energy would depend primarily on the quantity of radium present, on the degree of filtration and

on the time of exposure. But therapists are more concerned with the quantity of energy absorbed at a certain point in the tissue, and owing to the complex conditions which exist there is no simple or accurate way of relating the energy absorbed by the tissue to the energy emitted by the source. The energy absorbed by tissue from the electrons produced in it both by X rays and gamma rays can however be expressed in terms of the degree of ionization which they produce in air. Hence the present tendency is to favour an ionization method of measuring gamma-ray dosage and, in particular, for the sake of unification with X-ray dosage, to adopt the röntgen. The question of the suitability of the röntgen still however requires much attention. The International Committee for Radiological Units expressed the opinion at the Fourth International Congress of Radiology held at Zurich in July 1934 "that the ionization method of measuring gamma-ray dosage in röntgens promises to yield satisfactory results. It is further of the opinion that, before making a definite recommendation, the investigations already started in different countries should be continued and more data obtained".

At the time of writing this report, there are still insufficient data to lead to a complete solution of the problem and it will therefore only be possible to summarize the present state of the researches.[5–24]

Most success has attended the use of the air-wall chambers. It has been found that, owing to the effect of recoil electrons from the air and objects surrounding a chamber, the ionometric conditions are complex unless the wall thickness is at least equal to the electron range in the particular material. For air-wall chambers which satisfy this condition, the more recently published results give $8r \pm 10$ per cent for the equivalent of the Sievert unit dose, that is, of the gamma radiation received in 1 hour at 1 cm. from a point source of 1 mgm. of radium with a 0·5 mm. platinum filter. Such differences as occur are probably due to the inapplicability of the inverse square law of radiation as a result of using finite radium sources, finite measuring chambers and small radium-chamber distances.

Up to the present, there has been little consistency in the few results obtained with free-air chambers. The measured röntgen-equivalents of the Sievert dose vary from 2 to $4r$. Thus the readings with free-air chambers are very much smaller than those given by thick air-wall chambers. The reason appears to be that the experimental conditions in the case of thick air-wall chambers and of the free-air chambers so far employed are not comparable, probably due to the relative smallness of the free-air chambers. With thin air-wall chambers it is possible to obtain röntgen values of the order of $4r$ comparable with the free-air results, by using a well-defined beam of gamma radiation such as is essential with free-air chambers. As the experimental conditions in such cases are comparable, it would appear that there is a similarity of behaviour between free-air and air-wall chambers, which suggests that the röntgen is realizable in the gamma-ray region. More experimental work is, however, desirable before finally adopting the unit.

X-ray and gamma-ray protection. The *International Recommendations for X-ray and Radium Protection*[25] and the *Recommendations of the British X-ray and Radium Protection Committee*[26] were both revised in 1934. There is little to add regarding

X-ray protection, where such modifications as occur are mainly intended to keep pace with the use of X rays excited at increasingly higher voltages. In the realm of radium protection it was felt that more information was desirable, in view of the large quantities of radium now in use, particularly for teletherapy.

The X-ray protective recommendations are based on an experimentally determined "tolerance dose" of 0·2 röntgen per day (equivalent to continuous irradiation at a dosage rate of $10^{-5}r$ per sec.). In the case of the gamma rays from radium, there is not at present sufficient data to evaluate a tolerance dose. On the assumption, however, that the body is equally susceptible to X rays and gamma rays, a figure of 0·2 röntgen of gamma radiation per day has been tentatively adopted by the International and British Protection Committees and also by the Health Organization of the League of Nations[27].

Certain important factors in gamma-ray protection, such as remoteness and lead shielding have recently been studied at the National Physical Laboratory, particularly as regards the storage, manipulation and transport of radium. In addition, the radiation field has been determined for a 1-gramme radium bomb, which is typical of those employed in teletherapy. The investigation, which gives some idea of the doses received and of the desirable lead protection under typical working conditions, shows that the recommendations of the International and British Protection Committees provide a reliable basis for radium protection.

§ 4. RADIUM

By W. E. Perry, B.Sc.

For some time after its discovery in 1898, radium was obtained solely from Bohemian pitchblende, the yearly production being 1 or 2 gm. Later, several hundred grammes were produced from the carnotite minerals in Utah and Colorado. Uranium minerals were found in the Belgian Congo in 1913, but, owing to the war, the extraction of radium from this source was not carried out until 1922, since when, Belgium has become the premier radium producing country. Rich deposits of pitchblende were discovered in 1930 in the Great Bear Lake region of Canada[1] and extraction of radium from the ore in this locality is now being carried out at the rate of about 2 gm. per month.

The demand for radium arises mainly from its use for medical purposes, for which quantities ranging from 0·5 mg. to several grammes are required. It has also been employed in radiography, its small bulk and the superior penetrating power of the gamma rays giving it advantages over X rays in certain cases. Small quantities are also used in the form of luminous paints, although mesothorium is frequently the chief radioactive constituent of such materials.

The measurement of radium in quantities greater than 0·01 mg. is carried out by the gamma-ray method. This consists in comparing the gamma-ray ionization with that due to a standard containing a known quantity of radium. International Standards, consisting of weighed quantities of pure radium chloride, were set up at Paris and Vienna in 1912, and secondary standards, based on gamma-ray com-

parisons with one or both of these standards were issued to a number of countries. In 1934 the original International Standards were replaced by new standards. At the same time the British National Radium Standard, which had been in the custody of the National Physical Laboratory since 1912, was exchanged for a weighed standard. This is certified to consist of 20·50 mg. of anhydrous radium chloride, equivalent to 15·60 mg. of radium element. The radium salt is contained in a thin walled glass tube. It is part of a very pure specimen which, in a recent determination by Hönigschmid of the atomic weight of radium yielded the value 226·05. Inter-comparison, by the gamma-ray method, of the old and new British standards showed that the certified radium content of the old standard, after allowing for radioactive decay during the period 1912–34, was within the limits of experimental error (0·3 per cent) the same as the value based on the new standard.

Electrometers and electroscopes are still the standard instruments for the accurate measurement of radium. Summarized accounts of the various designs of ionization chambers and the experimental arrangements adopted have frequently been given[2]. In view of the need for rapid measurement, direct reading methods have received considerable attention, the following example[3] being typical. The ionization current due to the gamma rays is passed through a resistance of the order of 10^{10} ohms, consisting of a polished graphite streak on an insulating material. The steady potential between the ends of the resistance is measured by a Lindemann electrometer. The apparatus is made direct-reading by suitable adjustment of the sensitivity and calibration with a known quantity of radium, the accuracy claimed being about $\frac{1}{2}$ per cent.

Ordinary triodes, electrometer valves (valves having a high input resistance and operating at plate voltages of 4 to 10 V.) have been used successfully in the amplification of gamma-ray ionization currents, but this type of apparatus has not yet been adopted generally for routine radium measurements. The ionization currents are generally less than 10^{-11} A. for the preparations usually met with, so that to obtain a suitable deflection on a microammeter, an amplification of at least 10^6 is necessary. A two-stage amplifier[4], consisting of a screen grid valve operating at a plate voltage of 6 V., resistance-coupled to the second valve, has been described which yields a current of 1 μA. for an ionization current of 2×10^{-12} A. Such an amplifier in conjunction with a specially designed ionization chamber has given results comparable in accuracy with those obtained by a gold leaf electroscope. A valve amplifier was used in a test by a modification of Bothes method which showed that a sample of Canadian radium was free from mesothorium within the limits of experimental error[5] (0·2 per cent). A Geiger counter, highly sensitive to gamma rays, has been suggested as an instrument for detecting radioactive ores, etc., and its application to the measurement of the radioactivity of rocks has been investigated[7]. A simple, inexpensive and easily portable apparatus[8], incorporating a small neon glow lamp which acts in a manner similar to that of a Geiger counter, has been used for the measurements of ionizing radiations and for locating radium tubes and needles which are occasionally lost in hospitals. The discharges in the lamp are

made to operate a loud speaker or telephones so that an audible indication of the presence of the lost radium container is obtained.

Radium salts for use in medicine are generally enclosed in metal containers, such as needles, tubes, applicators etc. The metals used are chiefly platinum, silver, gold and monel metal, the thicknesses ranging from a fraction of 1 mm. to several millimetres. In determining the radium content, a correction is necessary to allow for the absorption of the gamma rays in the walls of the container. The correction may be calculated, using the appropriate absorption coefficient of the metal in question. Several factors, one of which is the scattering of gamma rays, cannot however be easily taken into account in the calculation and absorption corrections are more accurately obtained by direct measurement. The corrections vary with the experimental arrangement adopted and with the thickness of the lead filter which is usually interposed between the radium and the measuring instrument. For the routine measuring apparatus used at the National Physical Laboratory[9] they range from about 4 to 22 per cent for platinum containers of wall thickness 0·3 to 1·5 mm. Further corrections are necessary to allow for absorption in the radioactive material. This question has been investigated by F. Behounek[10], who measured the absorption coefficients of the gamma rays in a number of salts. An empirical relationship between the coefficient and the number of electrons per cm.^{3} of the material was deduced and used to obtain the salt absorption correction for mixtures of radium and barium chlorides.

§ 5. X-RAY SPECTROSCOPY

By G. Shearer, M.A., D.Sc.

One of the most interesting questions which arise in connexion with the measurements of X-ray wave-lengths is the expression of the wave-lengths in terms of the centimetre. When a crystal is used the wave-length is determined as a multiple of the grating constant of the crystal by means of the modified Bragg equation

$$n\lambda = 2d\,(1-\delta/\sin^2\theta)\sin\theta,$$

where λ is the wave-length, d the crystal spacing, θ the glancing angle for the nth order of reflection and $1-\delta$ is the refraction index of the crystal for the rays in question. The value of θ can be determined with a much greater accuracy than that with which d is known. For this reason it has been the custom to assume a value of d and to express wave-lengths in X units where 1 X.U. has the approximate value of 10^{-11} cm. The value of d for the cleavage face of calcite is taken as 3029·04 X.U. for the first order reflection at 18° C.

The value of d for calcite is given by the formula

$$d = \sqrt[3]{\frac{M}{2N\rho\,\phi\,(\alpha)}},$$

where M is the molecular weight of $CaCO_3$, N is Avogadro's constant, ρ the density and $\phi\,(\alpha)$ a numerical factor depending only on the rhombohedral angle of

calcite. The most uncertain factor in this expression is N, and the uncertainty in N is due to the fact that its determination involves a knowledge of the electronic charge e. The value usually adopted for e is that based on Millikan's oil drop method and Birge gives the value $4{\cdot}768 \times 10^{-10}$ e.s.u. as the most probable result of these experiments.

The discovery that ordinary ruled line gratings could be used to determine X-ray wave-lengths offered a means of determining such wave-lengths directly in terms of the centimetre. The early measurements gave values higher than those obtained by crystal methods, but as the degree of accuracy of such experiments was rather low, it was assumed that the discrepancy between the two sets of measurements arose from experimental errors in the grating values. As the technique of the method improved it was found, however, that this discrepancy did not diminish but that the grating values appeared to be consistently higher than the crystal values, the amount of the difference being approximately 0·2 per cent of the wave-length. During the past year several new and accurate determinations of wave-lengths by the grating method have been carried out notably by Bäcklin[1], Söderman[2] and Bearden[3] and these have confirmed the earlier work of Bäcklin, Purdom and Cork, Bearden and others. As very elaborate precautions were taken to avoid all known sources of error we would appear to be justified in taking these values as correct. If we do so, it is clear that they can be used in the crystal formula to determine N or e. When this is done the resulting value of e is $4{\cdot}805 \times 10^{-10}$ e.s.u. with an error of about $\pm 0{\cdot}003 \times 10^{-10}$ e.s.u. This value is well outside the error limits given by Millikan whose original determination was $4{\cdot}770 \pm 0{\cdot}004 \times 10^{-10}$ e.s.u.

Attempts have been made to explain the different values obtained by the two methods by attributing it to the use of a wrong value of the density in the formula from which d is calculated. The density of the crystal is obtained by macroscopic methods and it has been suggested that this value may not be the one which should be used to convert the volume of the unit cell of the crystal into the mass associated with that cell. For example, it has been pointed out that in crystal reflection experiments the surface layers of the crystal play an important part and the density in these layers might differ appreciably from that which characterized the crystal as a whole. Experimental investigations on this point by Allison and Williams show the hypothesis to be untenable. Zwicky's theory that the crystal structure is not perfect but that there is in every crystal a periodic block structure also suggests that the value used for the density may be in error. Here again the experimental evidence for crystals of widely varying density suggests that any error introduced in this way is much too small to account for the observed differences in wave-length. It would therefore appear that, in the present state of our knowledge, we must accept the wave-lengths as given by the line grating measurements as essentially correct and conclude that the most probable cause of the discrepancy between them and the crystal values is the uncertainty in the value of e. In this connexion it may be noted that Bearden has recently determined X-ray wave-lengths from measurements of the refractive index of quartz and diamond. Although it is possible that the formula used connecting the wave-length and the refractive index may not be

rigorously true it is interesting to find that his results agree with the grating results and lead to a value $4{\cdot}8036 \times 10^{-10}$ e.s.u. for e.

If we accept the grating measurements as correct, it follows that the X unit has the value $1{\cdot}0022 \times 10^{-11}$ cm. and that the grating constant for calcite must be correspondingly increased to give its value in the c.g.s. system. If the true value of e is $4{\cdot}805 \times 10^{-10}$ e.s.u. the value of Planck's constant h must also be increased since all determinations of this constant involve the value of the electronic charge. The corresponding value of h would be approximately $6{\cdot}625 \times 10^{-27}$ erg sec. It is of interest to note that these values agree within error limit with the theoretical equation $hc/2\pi e^2 = 137$ deduced by Eddington. In this equation the right-hand member is an integer which represents the number of degrees of freedom of a system of two particles in the space employed.

It must be admitted that it is extremely difficult to give reasons why the value obtained by Millikan should be in error. The only suggestion is one put forward by Shiba[4] who considers that the figure used by Millikan for the viscosity of air is the cause. Inserting a value which he considers more reliable, he obtains for e the value $4{\cdot}803 \times 10^{-10}$ e.s.u. and deduces $6{\cdot}624 \times 10^{-27}$ as the most probable value of h.

The electronic charge is not the only fundamental constant which presents difficulties when we attempt to assign to it an accurate numerical value. A somewhat similar difficulty appears to be associated with the quantity e/m. The best experimental value of this constant appears to be $1{\cdot}757 \times 10^{7}$ e.m.u. whereas the value determined by more indirect methods appears to be higher. Eddington has put forward a theory in which besides obtaining the equation $hc/2\pi e^2 = 137$, he gives a second equation

$$10x^2 - 136x + 1 = 0,$$

the ratio of whose roots should give the ratio of the mass of the proton to the mass of the electron. This equation gives $M_P/m = 1847{\cdot}6$ from which it can be deduced that e/m has the value $1{\cdot}7703 \times 10^{7}$ e.m.u., a value very considerably greater than that deduced from the latest spectroscopic or deflection experiments. Bond[5] has put forward an interesting suggestion in this connexion. He comments on the fact that when Eddington first calculated the value of $hc/2\pi e^2$ he obtained the result 136 which he amended later to 137 in order to take into account the degree of freedom associated with the indistinguishability of the particles. Bond suggests that in some or possibly all the methods used for the determination of e/m a similar effect has occurred and that what is determined experimentally is not e/m but $136/137$ times e/m. Eddington[6] is inclined to agree that some such thing may have happened and that a factor of $137/136$ neglected in current quantum theory will be introduced by the change in metric when we equate the space occupied by the indistinguishable particles of the quantum theory to the space occupied by the distinguishable parts of our measuring apparatus. Birge,[7] in commenting on Bond's suggestion and in emphasizing how accurately the two values of e/m are in the ratio of 136 to 137, points out that an almost similar ratio exists between the oil drop value of e and the value determined by X-ray methods!

In spite of the additional evidence which has been obtained in support of the grating wave-lengths it would be unwise at present to consider these values as definitely establishing the magnitude of the X unit and of the electronic charge. No doubt in the near future further evidence on these questions will be forthcoming.

In all branches of X-ray spectroscopy steady progress continues to be made both in the measurement of the wave-lengths of the various lines and in the interpretation of these measurements. The use of concave gratings by Siegbahn and his school has led to considerable advances in our knowledge of emission lines and absorption phenomena in the region 20–500 A.

An interesting series of experiments which have appeared during the year concerns the width of X-ray lines. After allowing for the width produced by experimental conditions, there remains a finite residue which is a real property of the line itself. X-ray lines are much broader than those of the ordinary optical spectrum. In the latter region the width of a line is probably less than 0·0001 per cent of the wave-length whereas in the case of the K lines of tungsten the width is as much as 0·1 per cent of the wave-length. Measurements of these widths have been made by Richtmyer[8] and others and are of considerable importance in determining the effective widths of the various energy levels in the atom. Thus it is found that the $K\alpha_1$, $K\beta_1$ and $K\gamma_1$ lines of tungsten have widths expressed in terms of volts of 43, 50 and 37 respectively. The widths in the L series are considerably smaller.

On the assumption that the width of a line is the sum of the widths of the two absorption levels concerned Richtmyer and his co-workers give a fairly complete scheme for the levels of gold. The K level has a width of 40 V., the L_I L_{II} L_{III} levels 8·7, 3·7 and 4·4 V., the M_I M_{II} M_{III} M_{IV} M_V widths of 15·5, 10·7, 12·1, 4·2 and 3·5 V., and so on. The $K\alpha_1$ line corresponding to the transition K–L_{II} would then have a width of 40 + 3·7 V., the $K\beta_1$ line corresponding to K–M_{II} a width of 50·7 V. Allowing for the difficulties associated with the measurements the scheme gives consistent results and further work on these lines should be of considerable interest.

§ 6. CRYSTAL STRUCTURE OF METALS AND INORGANIC SUBSTANCES

By W. A. Wood, M.Sc.

It is not possible to review adequately in the space available all recent advances in this subject. We select therefore one aspect which has attracted attention in the past year, that of the "real structure" of crystals which formed the subject of discussion at the International Conference on Physics held in London, October, 1934, partly under the auspices of this society.

Problem. The problem at issue arises in connexion with the discrepancy between the theoretical and observed strength of materials. The elastic constants and energy of the crystal lattice calculated by Born's theory show good agreement with experiment; yet, calculated on the same assumptions, the strength of crystals, or, the

force required to separate two parts across a crystal plane, is in disagreement. Thus Zwicky[1] showed that the strength of rock salt should be 200,000 gm./mm^2; actually it is about 450 gm./mm^2. The same discrepancy obtains in metallic crystals (Polanyi and Schmidt)[2]. Other physical properties subject to the same variability include electrical and thermal conductivities, diffusion, plasticity, dielectric and tensile strengths, properties termed by Smekal[3] structure-sensitive.

Evidence for flaw systems. The solutions put forward vary greatly in detail but in general agree in postulating some system of flaws which is regarded as an inherent and inevitable property distinguishing the real crystal from the mathematical model. Evidence for these flaws may be classed roughly as follows:

(i) If flaws do exist then the strength of a crystal is vitally affected. Thus early experiments of Griffith[4] showed that the strength of quartz and glass rods, protected from surface flaws, could be increased tenfold; also, the strength of rock salt measured in hot water was shown by Joffe[5] to reach 10,000 gm./mm^2, and values up to 20,000 gm./mm^2 are attained, as shown by Piatti[6], when rock salt is carefully stretched in salt-solution at room temperature; similarly Orowan[7] found that the strength of mica, clamped in such a way as to make the edges stressless and the characteristic edge cracks inoperative, increased from 30 to 300 kg./mm^2, a remarkably high value. Such experiments appear to indicate that the high values of the theoretical strength are correct and that the observed values are fortuitous.

(ii) Further evidence arises from attempts to interpret surface markings on crystals, etch figures or distributions of foreign atoms. These phenomena have long been familiar to crystallographers and metallurgists; but they have been studied in relation to the present problem by Zwicky[8], Goetz[9], Buerger[10] and Smekal. Thus Goetz photographed sets of lines on the basal cleavage plane of bismuth and claimed that they formed a regular triangular pattern of lines spaced at $1{\cdot}4 \times 10^{-4}$ cm., and were related to the triangular pits formed on etching. Dean and Koster[11] attribute a regularity to the characteristic tear lines of galena. Straumanis found that zinc and cadmium crystals deposited from the vapour phase formed a stepped arrangement of plates about $0{\cdot}8 \times 10^{-4}$ cm. thick. The etching pits in iron studied by Belaiew have been taken to indicate a block-like formation with a side of $0{\cdot}25 \times 10^{-4}$ cm. Many years ago a statistical regularity in the striae produced by translation gliding in metals was pointed out by Rosenhain. The interesting point about these observations is the value of periodicity claimed for the faults; this is nearly always of the order of $1\,\mu$.

(iii) The third line of approach is probably the most direct. The reflection of a parallel beam of X rays from a crystal rotated slowly through the reflecting position should take place within a few minutes of arc; actually the range of the reflecting position may extend from several minutes to several degrees even with crystals which appear optically perfect[12]. This appears to be irrefutable evidence that real crystals consist of independently reflecting units differing slightly in orientation from the mean crystallographic orientation of the parent crystal. Darwin, Bragg, and James concluded that the X-ray data on their rock salt crystals indicated a mosaic element of $0{\cdot}5 \times 10^{-4}$ cm. Further evidence arises from a comparison of

observed intensities of reflection of the X-ray spectra with the theoretical values formulated by Darwin and Ewald for a perfect and for a mosaic type of crystal. This has been done by Bragg, James, Bosanquet and others for sodium and potassium chloride and aluminium which appear definitely to agree with a mosaic structure, for diamond which apparently forms an almost perfect crystal, and for calcite which seems to be intermediary.

The trend of the evidence is in the direction of conceding to a crystal a series of flaws or discontinuities which appear to tend to a certain statistical size or distribution. Attempts have been made however to read into the results indications of definite systems of flaws and to advance formal theories which are very briefly referred to below.

Formal theories of the flaws. (i) *Superstructures* (*Zwicky*)[8]. Zwicky has formulated the theory that every *n*th plane in a crystal shows a departure from normality and that a large-scale lattice or superstructure is thereby superposed on the theoretical lattice. These abnormal planes are characterized according to his later papers by an excess in the number of atoms per unit area. Direct experimental evidence for the superstructures is adduced from observations on surface markings such as those described above, and from work by Johnson[13] who found diffraction effects in the reflection of atomic beams from crystals which he thought were due to abnormally large spacings. Indirect experimental evidence claimed in support of the theory are results by Goetz[14] on a difference in thermal expansion coefficients of bismuth obtained first on the macroscopic crystal, and, second, from X-ray data on the expansion of the lattice; and also the possible differences between the values of X-ray wave-lengths determined from crystal and from direct grating measurements. Then, by theory, Zwicky has sought to show that there is a decrease in the energy of a system on the formation of a secondary structure so that the latter is a more stable state; he starts from the point shown by Lennard-Jones and Dent that the (100) face of rock salt if set free as a separate plane of atoms would contract by about 6 per cent linearly, and assumes the effect might happen also for certain planes within a crystal, the sets of contracting planes forming a superstructure with a decrease in free energy. Exception has been taken however to these calculations by Pauling, Orowan[15] and others. Objections have been made also to the interpretation of absolute regularity in the Goetz and similar structures which unfortunately are not very easily resolved. A. H. Jay has found no difference in the expansion coefficients measured by the X-ray and the standard methods for the cases of bismuth, silver and quartz[16].

(ii) *Block structure* (*Smekal*). Smekal and his school consider a crystal is composed of very small blocks arising inevitably out of such accidents of growth as gaps, variations in orientation, impurities and so forth; these he terms primary flaws as distinct from secondary flaws which may follow upon later heat or mechanical treatment; the primary are in general submicroscopic but the secondary may attain easily observed sizes. A great number of interesting papers by Smekal and others have appeared on the detection of such flaws and on relations between the boundaries of the block system and such properties as diffusion between the

blocks, photoelectric coloration at flaw centres as well as the strength of crystals[21]. A criticism which arises is that whilst the existence and influence of internal flaws seems obvious, the reason why they should form any definite block-structure does not appear clear.

(iii) *Lineage structure (Buerger)*[17]. Buerger suggests that real crystals contain an internal structure of "lineages" or branches which radiate outwards from the original nucleus or nuclei, dividing again into sub-lineages until the whole volume of the crystal is solid. This theory differs from the preceding in that the boundaries of the lineages do not necessarily divide up the crystal into separate compartments. Such crystals are familiar as dendritic structures but Buerger claims that all crystals contain signs of this type of formation. He cites a number of cases which, observed through the microscope, appear like dendritic skeletons filled in by numerous branches. This theory at present is open to the same criticism as the above, that it gives no definite quantitative indications by which it could be tested.

(iv) *Statistical flaw-distribution (Griffith, Joffe)*. By this theory the weakness of crystals arises from the presence of flaws which pervade the crystal without reference to any special distribution as required by the secondary, block or lineage structures. Joffe considers the flaws become most effective when they are on the surface and tends to disregard the influence of small internal flaws. The early theory of Griffith attributed weakness to the concentration of applied stress at minute cracks or crevices; this local stress might exceed the mean stress and approach the theoretical strength, when a crack would grow and propagate itself through the material. The size of the cracks required in the first case are however too large on this simple theory; but Orowan has pointed out that this objection is easily got over by taking into account the process of slip in the neighbourhood of a flaw.

The general position appears to be that the Zwicky structure is viewed with disfavour, and that the block and lineage structures involve formations which are too inflexible to accord with all observations. The strength of a crystal is also a function of the degree of plastic strain which it exhibits before fracture; usually this factor is overlooked. There is still need for an explanation of properties which remain when accidental variations are allowed for, in particular of the nature and mechanism of plastic deformation. Definite progress has recently been made in this direction by G. I. Taylor[18], Polanyi[19] and Orowan[20]. The ideas of the former lead to quantitative results and are summarized below.

(v) *Taylor's theory of plasticity*. During deformation of a crystal by slip, the translation preserves the same crystallographic direction even after large amounts of shear. This indicates that one plane of atoms moves relative to its neighbour in jumps of one atomic separation so that the crystal lattice is continuously re-formed and essentially unaltered.

The first slip however evokes a resistance to further shear, so that the process of slip must produce some difference in the lattice. This is explained by imagining the slip to take place not over the whole of a plane at once but only over limited areas, so that the strained crystal would contain lines of dislocation separating the portions which have moved from those which have not. The observed external

strain would then be a combination of these limited movements; the presence of the dislocations would give rise to elastic stresses in the material.

The force required to start the slip is small in practice compared with that necessary in theory to move bodily one plane of atoms over the next. The potential barrier normally surrounding an atom must therefore be lowered to allow it to pass into a neighbouring position. This is got over by showing that at a dislocation of the type pictured such a reduction would occur. The centres of dislocation could be regarded therefore as moving along the slip plane and permitting the jumps of successive atoms. In a perfect crystal under no external stress the dislocations are supposed to be freely mobile to an extent depending on temperature; when stress is applied in excess of a critical value depending on the temperature an overriding tendency would propel the dislocations along the slip plane in the direction of the resolved stress. A positive dislocation is regarded as one where $N+1$ atoms in the plane over the centre of the dislocation face N atoms in the plane below; and a negative one as the reverse. Taylor estimates the stress due to such dislocations several atomic distances away and shows that under an applied stress the positive ones repel and the negative attract until equilibrium is reached. The process of shear is then regarded as the generation and separation of positive and negative dislocations under a stress S to a mean distance L. If the separation of a number of pairs produces a strain s, then Taylor shows that $S/\sqrt{s}=\mu K\sqrt{(\lambda/L)}$, where μ is the coefficient of rigidity, λ the atomic separation and K a constant depending on the arrangement of the centres of dislocation.

This result is notable as it is the first quantitative relationship derived between shear and resistance to shear from a general conception of the mechanism of slip. Taylor shows that the formula fits reasonably the observed parabolic (S, s) curves for aluminium crystals (Bons and Schmidt) and the observations of Joffe and Piatti on rock salt.

Taylor compares his estimate values of L with the periodicities put forward in the various systems of faulting and points out that it is of the same order, 1μ. The theory therefore requires only that at a mean distance L there should be some condition opaque to the passage of the dislocations; it does not require any particular distribution of the Zwicky or Smekal type, and it assigns a definite function to the faults in relation to the strength of the crystal. This and the related work of Polanyi and Orowan represent definite progress toward the solution of the original problem.

§ 7. CRYSTAL STRUCTURE OF ORGANIC AND BIOLOGICAL SUBSTANCES

By J. Thewlis, M.Sc.

Most organic crystals possess low symmetry. They are composed largely of hydrogen atoms, which give no X-ray reflection, and of carbon, nitrogen and oxygen atoms, the reflections from which are practically indistinguishable. Consequently complete determinations of structure have been made for very few

compounds. The determinations of structure in other cases have been restricted to establishing the crystal type, size of unit cell, molecular symmetry, etc. It is possible, however, to get a great deal of information about the molecular structure of a substance from the determination of these factors, by studying the effect of adding, subtracting or substituting comparatively heavy atoms in the molecule[1]. The positions of these atoms in the unit cell are relatively easy to discover, since such atoms are characterized by a high scattering power. Other methods have also been found useful in this work. For example, information provided by Raman spectra, piezo-electric measurements, optical and magnetic measurements, or by a knowledge of cleavage, has been of great service in the elucidation of organic structures.

The structures of so many organic compounds have now been investigated that it is impossible, in a short article such as this, to do more than touch on a few aspects of the subject and possibly to give some account of the general trend of the work. Of necessity many omissions must occur.

As a result of the many investigations which have been carried out it has been found that the modern methods of X-ray analysis have to a great extent verified the older conceptions of organic structure put forward by chemists. A striking example is that of the benzene ring. Not only has this ring been found to have an actual existence, but its shape has been determined, and all the evidence goes to show that it takes the form of a flat regular hexagon. The structure of benzene itself has been studied[2], but the best evidence that the benzene ring is flat comes from a study of the more complex aromatic structures, particularly that of durene[3].

The distance from centre to centre of neighbouring atoms in the benzene hexagon is 1·41 A. and this value is the same in all compounds yet examined in which the benzene hexagon occurs. This distance is maintained, for example, in the polyphenyl structures, and in addition a distance of 1·48 A. between the hexagons is noted[4]. The study of condensed ring systems in general shows that they form flat rigid molecules, which with an increasing number of rings tend to pack more and more into parallel sheets, approximating to the structure of graphite.

Much work has been carried out on the structures of long-chain aliphatic compounds such as the paraffins and fatty acids[5]. They are found to be based on zig-zag chains of carbon atoms, the spacing increasing regularly with the number of carbon atoms in the chain, for any one series of compounds. The crystalline arrangement explains the observed properties of the compounds, such as the differences in physical properties of compounds with odd and even numbers of carbon atoms.

A substance on which many investigations have been made is cellulose[6]. It consists of polymerized chains of anhydro-β-glucose, and recent work indicates that the quasi-planar rings of the glucose residues are parallel to the (002) planes[7]. From a study of the dinitrocelluloses it appears that the NO_3 groups are distributed at random along the molecular chains, but in trinitrocellulose these groups are probably arranged at regular intervals along the chains[8].

The sugars have also been studied, but it has not been possible to do much

more than establish the space-groups and unit cells. What evidence there is, however, points to the existence of an almost plane ring[9].

It has been remarked above that the benzene ring is characterized in various compounds by a constant distance of 1·41 A. between neighbouring carbon atoms. Another subject of measurement is the angle between bonds. In the benzene ring, of course, the angle is 120°. The values of these interatomic distances and angles seem to depend only on the nature of the atoms and the degree of their association, i.e. whether they are bound by single, double, or treble bonds, and not on the structure in which they occur. Consequently a knowledge of their values is of great help in the determination of structures.

Another value which tends to be constant, and hence very useful, is the distance between neighbouring atoms in different molecules. As might be expected, since the forces between molecules are comparatively small it is larger than the distance between neighbouring atoms of the same molecule. It actually averages about 3·5 A.

Values for interatomic distances can only be obtained from structures which have been completely analysed, and it may be of interest to give a list of organic structures of which complete analyses are available. These substances are: urea[10], thiourea[11], hexamethylbenzene[12], durene[3], naphthalene[13], anthracene[14], *p*-diphenylbenzene[4] and chrysene[15]. In these substances the aromatic bond is 1·41 A., as already mentioned, and the aliphatic bond is about 1·54 A. The value of the aliphatic-aromatic bond found in hexamethylbenzene and durene is 1·48 and 1·47 A. respectively, i.e. it is intermediate between those of the aliphatic and aromatic bonds.

We come now to the biological applications of X-ray diffraction, and here again only a general account can be attempted.

Arising from the investigation of the crystal structure of calciferol (vitamin D)[16] and its related compounds, and of cholesterol and ergosterol, a host of similar substances has been examined. It has been found that the sterol molecule forms a double layer in the crystal, similar to that of long-chain alcohols and a new sterol formula has been propounded which accounts for all the features occurring in the crystal analysis[17]. This and other work has shown that the structures of a number of groups of biochemical compounds, including the sterols, the bile acids and the male and female sex hormones, are based upon a common condensed ring system.

Vitamins B_1, B_4 and C (*l*-ascorbic acid) have been examined and a structural formula for the latter has been established[18]. The polyenes, which include carotene and vitamin A, have also been examined[19].

A structure which has only recently been examined is that of insulin[20]. The shape of the molecule is approximately an oblate spheroid of diameters 44 and 30 A. As a result of this work a structure for pepsin has also been proposed.

During the past few years an enormous amount of work has been done on the proteins. They are found to consist of polypeptide chains and can exist in the contracted (α) or expanded (β) form[21]. They have characteristic thicknesses and widths. The thickness is about the same for all proteins and is about 4·65 A., but the width, which corresponds to bonding between the side chains, varies with their

length and is 4·5–6·1 A. in silk fibrin and as much as 9·8 A. in keratin. The structure of the proteins has been found to be capable of explaining physical properties such as the stretching of hair and the adsorption of water. In particular much work has been done on the relations between the molecular structure and elastic properties of hair keratin[22].

An interesting point which has been brought out in this work is that the old classification of feathers and scales with animal hairs etc. under the heading of keratin is incorrect[23]. From the X-ray patterns it is clear that an affinity exists between reptiles and birds, the X-ray photographs of their epidermal proteins being similar to each other and different from those of mammals.

The structure of bone has also been investigated[24]. The inorganic constituent has the apatite structure and is stated to correspond to the mineral dahllite. In fossilized bone, however, fluor-apatite has been found.

The organic lattice in bone shows typical fibre-structure of a spiral nature, but this changes to simple fibre-structure on decalcification. As the bone gets older its organic content becomes less. In extreme age the apatite pattern obtained from bone becomes weaker, i.e. it would appear that the bone is tending to become amorphous.

Teeth are similar to bone in that their inorganic constituent has an apatite structure[25,26]. Its precise nature has not however been established. It has been said to be hydroxy-apatite[27] but the evidence does not appear to be conclusive.

The enamel of teeth is marked by a very well defined preferential orientation of the apatite crystallites, and the fibre-axis varies in its relation to the enamel surface in man and animals[28]. Marked variations in the degree of preferential orientation have been noted even in the same tooth but the exact significance of them is not yet certain[29]. In contrast with enamel, dentine is characterized by a random arrangement of apatite crystallites[26].

Other portions of the body also give X-ray diffraction patterns which are similar to that of apatite. These include the calcific deposit from the aorta wall, and stones from the prostate and lymphatic glands[30,31].

As might be expected, the X-ray diagram of gall-stone is found to agree with that of cholesterin. The diagram of pancreas-stone is also very similar[30]. Stones from the kidney, bladder and ureter have been examined but beyond the determination of their probable lattice sizes, and the fact that they consist of large crystals, little definite information has been obtained[30,31,32].

From the early work it was concluded that unstretched muscle was amorphous, but more recent work seems to show that muscle exhibits a fibre-structure even in the unstretched state[33,34]. Stretching of the muscle has the effect of making this fibre-structure (which is due to myosin) appear much more pronounced.

Marked fibre-structure is also found in sinew, whereas ligaments show a much less well-defined fibre-structure[35]. The separate constituents of nerve have been examined, and here again fibre-structure has been found[34,36]. An attempt has also been made to detect differences between the X-ray photographs of stimulated and non-stimulated nerves but this has met with no success[34].

Among many other results arising from the X-ray examination of ocular tissues it has been found that the lens of the eye shows no sign of being crystalline[37]. The lens covering on the other hand shows fibre-structure and the arrangement of the micelles has been determined. The cornea and sclerotic have been found to give evidence of a crystalline structure. The effect of cataract has also been studied.

Did space permit, much other work might be described which, like most of the investigations dealt with in this section, lies on the border lines of physics, chemistry and biology. Much of this work is concerned, not so much with the measurement of small units of pattern, as with the determination of larger details. It represents the beginning of an attempt to understand the relations between organic and biological bodies, and their constituent atoms and molecules; an attempt which seems to promise a fair measure of ultimate success.

REFERENCES

§ 1. Production of X rays

(1) Kaye and Bell. *Brit. J. Radiol.* **7**, 476 (1934).
(2) Bell. *Brit. J. Radiol.* **7**, 654 (1934); **8**, 218 (1935).
(3) Kaye and Bell. *J. Instn elect. Engrs*, **73**, 457 (1933).
(4) Thorrus. *Acta Radiol.*, Stockh., **14**, 615 (1933).
(5) Roberts. *Brit. J. Radiol.* **7**, 685 (1934).
(6) Niemann and Zeyss. *Fortschr. Röntgenstr.* **45**, 596 (1932).
(7) Taylor. *Bur. Stand. J. Res.*, Wash., **9**, 333 and 561 (1932).
(8) Gaertner. *Z. tech. Phys.* **15**, 59 (1934).
(9) Bouwers. *Z. tech. Phys.* **8**, 271 (1927).
(10) Müller. *Proc. roy. Soc.* A, **117**, 30 (1927).
(11) Bouwers. *Physica*, Eindhoven, **10**, 125 (1930).
(12) Müller. *Proc. roy. Soc.* A, **125**, 507 (1929).
(13) Clay. *Proc. phys. Soc.* **46**, 703 (1934).

§ 2. Ionization in gases, absorption and scattering of X rays

(1) Fowler and Gibson. *Phys. Rev.* **46**, 1075 (1934).
(2) Lawrence and Livingston. *Phys. Rev.* **45**, 608 (1934).
(3) Beeck. *Ann. Phys.*, Lpz., **19**, 121 (1934).
(4) Beeck and Wayland. *Ann. Phys.*, Lpz., **19**, 129 (1934).
(5) Gaertner. *Ann. Phys.*, Lpz., **21**, 564 (1934).
(6) Hahn. *Phys. Rev.* **46**, 149 (1934).
(7) Grosskurth. *Ann. Phys.*, Lpz., **20**, 197 (1934).
(8) Kustner. *Z. Phys.* **70**, 324 (1931).
(9) *Reports on Progress in Physics*, p. 233, Physical Society (1934).
(10) White. *Phys. Rev.* **46**, 865 (1934).
(11) Carr. *Phys. Rev.* **46**, 92 (1934).
(12) Williams. *Phys. Rev.* **46**, 83 (1934).
(13) Harvey, Williams and Jauncey. *Phys. Rev.* **46**, 365 (1934).
(14) Wentzel. *Z. Phys.* **43**, 1 and 779 (1927).
(15) Waller and Hartree. *Proc. roy. Soc.* A, **124**, 119 (1929).
(16) Jauncey and Claus. *Phys. Rev.* **46**, 941 (1934).
(17) Harvey. *Phys. Rev.* **46**, 441 (1934).
(18) Debye. *J. Math. Phys.* **4**, 133 (1925); *Phys. Z.* **28**, 135 (1927).

§ 3. X-RAY AND GAMMA-RAY QUANTITY OR DOSE. PROTECTION

(1) JAEGER. *Phys. Z.* **35**, 665 (1934).
(2) TAYLOR and STONEBURNER. *Radiology*, **23**, 22 (1934).
(3) MAYNEORD and ROBERTS. *Brit. J. Radiol.* **7**, 158 (1934).
(4) JAEGER. *Phys. Z.* **35**, 841 (1934).
(5) GLASSER, PORTMANN and SEITZ. *Amer. J. Roentgenol.* **20**, 505 (1928).
(6) KESSLER and SLUYS. *Strahlentherapie*, **31**, 771 (1929).
(7) JONA. *Strahlentherapie*, **32**, 775 (1929).
(8) NEEFF. *Strahlentherapie*, **33**, 253 (1929).
(9) BRUZAU. *Ann. Phys.*, Paris, **11**, 5 (1929).
(10) GLASSER and MAUTZ. *Strahlentherapie*, **34**, 845 (1929).
(11) BRAUN. *Strahlentherapie*, **38**, 11 (1930).
(12) FAILLA and HENSHAW. *Radiology*, **17**, 1 (1931).
(13) SIEVERT. *Acta Radiol.*, Stockh., **12**, 300 (1931).
(14) ALBRECHT. *Strahlentherapie*, **42**, 328 (1931); **45**, 365 (1932).
(15) MAYNEORD. *Brit. J. Radiol.* **4**, 693 (1931).
(16) LAURITSEN. *Amer. J. Roentgenol.* **30**, 380, 529 (1933).
(17) GOLDHABER and GRIFFITH. *Brit. J. Radiol.* **6**, 656 (1933).
(18) JAEGER. *Phys. Z.* **35**, 273 (1934).
(19) MALLET. *Arch. Élect. méd.* **598**, 247 (1934).
(20) FRIEDRICH and others. *Strahlentherapie*, **51**, 1 (1934).
(21) SIEVERT. *Acta Radiol.*, Stockh., **15**, 193 (1934).
(22) NEEFF. *Strahlentherapie*, **51**, 650 (1934).
(23) GLASSER. *Strahlentherapie*, **52**, 137 (1935).
(24) KELLER. *Strahlentherapie*, **52**, 403 (1935).
(25) *International Recommendations for X-ray and Radium Protection* (Revised, 1934).
(26) *Recommendations of the British X-ray and Radium Protection Committee* (Fourth revised report, 1934).
(27) *Protective measures against dangers resulting from the use of Radium, Röntgen and Ultra-violet Rays.* League of Nations Publications, **3**; *Health*, **3**, 9 (1931).
(28) KAYE, BELL and BINKS. *Brit. J. Radiol.* **8**, 6 (1935).

§ 4. RADIUM

(1) SPENCE. *Canadian Mines Department Publication*, No. 727, pp. 55–92 (1931).
(2) KOHLRAUSCH. Radioaktivität. *Handbuch der Experimental Physik* (1928).
(3) GRIMMETT. *Proc. phys. Soc.* **44**, 445–50 (1932); **45**, 117–119 (1932).
(4) CURTISS. *Bur. Stand. J. Res.*, Wash., **10**, 583–89 (1933); **12**, 164–72 (1934).
(5) LAURENCE and FRIEND. *Canad. J. Res.* **10**, 332–4 (1934).
(6) SHRUM and SMITH. *Canad. J. Res.* **11**, 652–7 (1934).
(7) VOIGT. *Phys. Z.* **34**, 79–82 (1933).
(8) CHALMERS. *Brit. J. Radiol.* **7**, 755–63 (1934).
(9) KAYE, ASTON and PERRY. *Brit. J. Radiol.* **7**, 540–6 (1934).
(10) BEHOUNEK. *Z. Phys.* **92**, 7–8, 533–42 (1934).

§ 5. X-RAY SPECTROSCOPY

(1) BÄCKLIN. *Nature*, Lond., **135**, 32 (1935).
(2) SÖDERMAN. *Nature*, Lond., **135**, 67 (1935).
(3) BEARDEN. *Phys. Rev.* **46**, 759 (1934); **47**, 883 (1935).
(4) SHIBA. *Sci. Pap. Inst. phys. chem. Res.*, Tokyo, **19**, 97 (1932).
(5) BOND. *Nature*, Lond., **133**, 327 (1934).
(6) EDDINGTON. *Nature*, Lond., **133**, 907 (1934).
(7) BIRGE. *Nature*, Lond., **133**, 648 (1934).
(8) RICHTMYER, BARNES and BAMBERG. *Phys. Rev.* **46**, 843 (1934).

§ 6. Crystal structure of metals and inorganic substances

(1) Zwicky, F. *Phys. Z.* **25**, 223 (1924).
(2) Polanyi, M. and Schmidt, E. *Naturwissenschaften*, **17**, 301 (1929).
(3) Smekal, A. *Z. Kristallogr.* **89**, 386 (1934) (and for other references).
(4) Griffith, A. *Philos. Trans.* A, **221**, 180 (1921).
(5) Joffe, A. *The Physics of Crystals*. New York, 1928.
(6) Piatti, L. *Z. Phys.* **77**, 401 (1932).
(7) Orowan, E. *Z. Phys.* **82**, 239 (1933).
(8) Zwicky, F. *Phys. Rev.* **40**, 63 (1932).
(9) Goetz, A. *Proc. Nat. acad. Sci.*, Wash., **16**, 99 (1930).
(10) Buerger, M. J. *Z. Kristallogr.* **89**, 195 (1934) (and for other references).
(11) Dean, R. S. and Koster, J. *Studies in Mineral Physics*. U.S. Bur. of Mines, p. 17 (1933).
(12) James, R. W. *Z. Kristallogr.* **89**, 295 (1934) (and for other references). (Cf. M. Renninger, *Naturwissenschaften*, **22**, 334 (1934).)
(13) Johnson, T. H. *J. Franklin Inst.* **212**, 507 (1931).
(14) Goetz, A. and Hergenrother, R. C. *Phys. Rev.* **40**, 643 (1932).
(15) Orowan, E. *Z. Phys.* **89**, 774 (1934).
(16) Jay, A. H. *Proc. roy. Soc.* **143**, 465 (1934); *Z. Kristallogr.* **89**, 282 (1934).
(17) Buerger, M. J. *Z. Kristallogr.* **89**, 195 (1934).
(18) Taylor, G. I. *Proc. roy. Soc.* **145**, 362, 368, 405 (1934).
(19) Polanyi, M. *Z. Phys.* **89**, 660 (1934).
(20) Orowan, E. *Z. Phys.* **89**, 314 (1934).
(21) Smekal. "Structure-sensitive properties of crystals", *Handbuch der Physik*, 2nd ed. (Berlin, 1933).

§ 7. Crystal structure of organic and biological substances

(1) For pioneer work in this field see the papers of J. D. Bernal.
(2) Cox. *Proc. roy. Soc.* A, **135**, 491 (1932).
(3) Robertson. *Proc. roy. Soc.* A, **141**, 594 (1933).
(4) Pickett. *Proc. roy. Soc.* A, **142**, 333 (1933).
(5) Müller. *Proc. roy. Soc.* A, **138**, 514 (1933).
Müller. *Helv. chim. Acta*, **16**, 155 (1933).
(6) Meyer and Mark. *Der Aufbau der hochpolymeren organischen Naturstoffe* (Leipzig, 1930).
(7) Mathieu. *C.R. Acad. Sci.*, Paris, **192**, 1434 (1934).
(8) —— *C.R. Acad. Sci.*, Paris, **200**, 143 (1935).
—— *C.R. Acad. Sci.*, Paris, **200**, 401 (1935).
(9) Cox. *J. chem. Soc.* **138**, 2535 (1932).
—— *Trans. Faraday Soc.* **29**, 90 (1933).
(10) Wyckoff. *Z. Kristallogr.* **81**, 102 (1931).
Wyckoff and Corey. *Z. Kristallogr.* **89**, 462 (1934).
(11) —— —— *Z. Kristallogr.* **81**, 386 (1932).
(12) Lonsdale. *Proc. roy. Soc.* A, **123**, 494 (1929).
(13) Robertson. *Proc. roy. Soc.* A, **142**, 674 (1933).
(14) —— *Z. Kristallogr.* **84**, 321 (1933).
—— *Proc. roy. Soc.* A, **140**, 79 (1933).
(15) Iball. *Proc. roy. Soc.* A, **146**, 140 (1934).
(16) Bernal. *Nature*, Lond., **129**, 277 (1932).
(17) Rosenheim and King. *Chemistry and Industry*, **51**, 464 (1932).
(18) Cox. *Nature*, Lond., **130**, 205 (1932).
(19) Waldmann and Brandenberger. *Z. Kristallogr.* **82**, 77 (1932).
(20) Crowfoot. *Nature*, Lond., **135**, 591 (1935).

(21) Astbury. *Fundamentals of Fibre-Structure* (Oxford, 1933).
—— *Kolloidzschr.* **69**, 340 (1934).
(22) Astbury and Woods. *Philos. Trans.* **232**, 333 (1933).
(23) Marwick. *J. Text. Sci.* **4**, 31 (1931).
(24) Henschen, Straumann and Bucher. *Dtsch. Z. Chir.* **236**, 485 (1932).
(25) Roseberry, Hastings and Morse. *J. biol. Chem.* **90**, 395 (1931).
(26) Thewlis. *Brit. J. Radiol.* **5**, 353 (1932).
—— *Brit. Dental J.* **53**, 655 (1932).
Bale, Hodge and Warren. *Amer. J. Roentgenol.* **33**, 369 (1934).
(27) Klement and Trömel. *Z. phys. Chem.* **213**, 263 (1932).
(28) Thewlis. *Phil. Mag.* **19**, 291 (1935).
(29) —— *Brit. Dental J.* **57**, 457 (1934).
(30) Saupe and Klötzer. *Fortschr. Röntgenstr.* **47**, 352 (1933).
(31) Worschitz and Hermann. *Fortschr. Röntgenstr.* **50**, 3, 305 (1934).
(32) Burgers. *Fortschr. Röntgenstr.* **48**, 228 (1933).
(33) Boehm and Weber. *Kolloidzschr.* **61**, 269 (1932).
(34) Boehm. *Kolloidzschr.* **62**, 22 (1933).
(35) Saupe. *Fortschr. Röntgenstr.* **44**, 204 (1931).
(36) Handovsky. *Kolloidzschr.* **62**, 21 (1933).
(37) Hertel. *Arch. Augenheilk.* **107**, 259 (1933).

SPECTROSCOPY

Edited by HERBERT DINGLE, D.Sc., A.R.C.S., D.I.C.

§ 1. THE DETERMINATION OF WAVE-LENGTH STANDARDS

By C. V. JACKSON

The primary standard. Two new absolute determinations of the wave-length of the red line of cadmium were published in the course of the last year. Sears and Barrell[1] published their final value $Cd_R = 6438{\cdot}4708$ A. in normal air*. The work was carried out at the National Physical Laboratory by the method of Fabry and Perot. In many details they introduced great improvements over the original work of Fabry and Perot. The type of étalon used in this research is extremely satisfactory and moreover may be exhausted. This rendered possible the absolute determination of the wave-length of the red line of cadmium *in vacuo*.

The other determination was carried out at the Technische Reichsanstalt by Kösters and Lampe[2], who found the value 6438·4672 A. for the wave-length of Cd_R in normal air. They used a method based on the interferometer of Kösters, which differs considerably from the methods of Michelson and of Fabry and Perot.

The discrepancy of 0·0036 A. (or 1 part in 2 million) between these two recent determinations appears at first sight to be unduly large in view of the fact that the experimental error in evaluating the length of an end-gauge in terms of light waves is of the order of 1 part in 10 million, either by the method used by Sears and Barrell or by that used by Kösters and Lampe. The explanation of the large discrepancy lies in the difficulty of measuring an end-gauge against a line standard and in the fact that it is not possible to compare two line standards with an accuracy much greater than 1 part in 5 million. The line standards used by the investigators were copies of the international prototype, and it is therefore easy to understand how the discrepancy of 1 part in 2 million between the results of the two observers arose.

The source of the red line used by Sears and Barrell differed slightly from the original lamp of Michelson, in that it was filled with air at a pressure of 1 mm. Hg, instead of being highly evacuated. As was discovered by Pérard, this has the effect of improving the sharpness of the radiation and lengthening the life of the lamp. Sears and Barrell also made an absolute determination of the wave-length of the red line as emitted by the new Osram cadmium lamp. The result differed from that found with the Michelson lamp by only 0·0001 A. These observers[3] have also compared the wave-lengths of the red line emitted by this type of lamp with that emitted by the Michelson lamp. They came to the conclusion that when the former lamp is excited with a current not exceeding 1·1 A. the visibility of the fringes is as good as that given by the air-filled Michelson lamp, while the wave-length is identical within their experimental error (1 part in 16 million). This type of lamp

* The phrase "normal air" refers to dry air containing 0·03 per cent carbon dioxide at 760 mm. pressure and at a temperature of 15° C.

has also been investigated by Pérard[4] who came to a similar conclusion on the visibility of the fringes but gave no data on the wave-length. He also investigated a lamp which was modified by having the central part of the inner tube constricted. He found that if the current did not exceed 1·1 A. the sharpness of the fringes was as great as that of the fringes given by the ordinary type of Osram lamp, but that with a current of 2 A. the red line was reversed. The advantage of the lamp with the constriction was its greater intrinsic brilliancy.

Iron standards. There was no new interferometric investigation on the iron arc spectrum published in 1934, but it is worthy of note that a photographic map of the infra-red part of the spectrum (λ8800 to λ10250) has been published by Dingle[10].

Neon standards. An important interferometric investigation of the wave-lengths of the neon lines has been made by Meggers and Humphreys[5]. The wave-lengths of ninety of the faint lines in the visible part of the spectrum were measured to $\pm$0·001 A., using as standards the means of the 8-figure wave-lengths given by the Bureau of Standards[6] and C. V. Jackson[7], while 8-figure wave-lengths are given for neon lines previously adopted by the International Astronomical Union as secondary standards. These were measured against the primary standard, using interferometer plates coated with silver, and gaps of 25, 35 and 43 mm. The agreement with the earlier observations of the Bureau of Standards and of Jackson is very satisfactory: the systematic difference is well under 0·0001 A. The results of the three sets of measurements are given below, together with the I.A. values of 1922. The numbers in parentheses after the wave-lengths of Meggers and Humphreys and Jackson refer to the number of observations made. It should be emphasized that all these wave-lengths apply to observations made with étalons of resolving power not less than about 300,000 and separation not greater than about 4 cm. With

Table 1. Wave-lengths of the bright lines of neon.

I.A. 1922	B.S.	Jackson	Meggers and Humphreys
5852·488	·4880	·4876 (28)	·4878 (8)
5881·896	·8954	·8948 (28)	·8950 (8)
5944·834	·8343	·8343 (28)	·8340 (8)
5975·534	·5339	·5340 (24)	·5343 (8)
6029·998	·9970	·9973 (24)	·9968 (8)
6074·338	·3377	·3377 (28)	·3376 (8)
6096·163	·1630	·1630 (28)	·1630 (8)
6143·062	·0624	·0620 (28)	·0627 (7)
6163·594	·5937	·5941 (28)	·5937 (8)
6217·280	·2811	·2814 (28)	·2812 (8)
6266·495	·4950	·4949 (28)	·4952 (8)
6304·789	·7890	·7893 (28)	·7893 (8)
6334·428	·4280	·4280 (28)	·4276 (8)
6382·991	·9913	·9915 (28)	·9914 (7)
6506·528	·5278	·5280 (28)	·5277 (7)
6532·882	·8826	·8824 (23)	·8824 (8)
6598·953	·9528	·9530 (23)	·9528 (8)
6678·276	·2760	·2766 (11)	·2766 (8)
6717·042	·0427	·0427 (2)	·0430 (8)
7032·412	·4130	—	·4125 (5)

lower resolving powers the wave-lengths are shifted to the violet by blending of the main lines with those due to the rarer isotope of neon.

In the same communication, Meggers and Humphreys also give wave-lengths to eight figures of twenty infra-red neon lines.

Krypton standards. Meggers and Humphreys[5] also present new interferometer wave-lengths of 60 krypton lines. These were measured partly against the primary standard and partly against the B.S.-Jackson neon standards. In the violet and visible part of the spectrum the agreement of these new 8-figure wave-lengths with the earlier ones of Humphreys[8] and Jackson[9] is very good. In the case of the infra-red lines, however, there is a serious systematic difference between the data of Humphreys and those of Meggers and Humphreys (H. − M. & H. = +0·0015 A.). It is difficult to understand the cause of this discrepancy, in view of the fact that the systematic difference of the krypton wave-length measurements by Meggers and Humphreys against cadmium in one set and neon in the other is only 1 part in 50 million.

Argon. Meggers and Humphreys[5] give interferometer measurements of 87 argon lines covering the range λ3948 to λ10470. The lines are well distributed and mostly given to eight figures and have been measured both against the primary standard and against the B.S.-Jackson neon standards. The systematic difference between the measurements against Cd and Ne in this case is again less than 1 part in 50 million, which indicates that with étalons of 2½–4 cm. thickness the neon scale gives results practically identical with those given by the primary standard.

It is pointed out that lines of argon are considerably sharper than those of neon on account of the higher atomic weight of argon, and that the satellites due to the lighter isotope are so faint that it is impossible to observe them. This is not surprising, as the ratio of abundance of the two isotopes (40 and 36) of argon is 160 to 1. Since the isotopes are both of even atomic number there is no hyperfine-structure due to nuclear spin. For this reason, and because of the relative sharpness of the argon lines and their distribution throughout the visible and infra-red part of the spectrum, it seems that the argon spectrum should be the most useful of all the rare gas spectra for providing wave-length standards. It is undoubtedly desirable that further investigations of the wave-lengths of this spectrum be carried out.

Xenon and helium. Meggers and Humphreys[5] have given interferometer measurements of 130 lines of xenon, but, as they point out, these are of very little use as standards on account of the relatively large abundance of the odd isotopes of xenon and consequent high intensity of the hyperfine-structure components due to nuclear spin.

Meggers and Humphreys[5] have also measured the wave-lengths of the three strong lines of helium in the infra-red, but these are of little value as wave-length standards, on account of their great breadth, due to the low atomic weight of helium.

§ 2. ISOTOPE EFFECTS IN THE SPECTRA OF DIATOMIC HYDRIDES

By R. W. B. Pearse, A.R.C.S., Ph.D.

Recently, there has been great activity in producing and analysing spectra of diatomic deuterides and in comparing them with the spectra of the corresponding hydrides. The elementary theory of isotopic displacements, which considers only the effects of the change in value of the reduced mass, μ, in the usual expressions for the vibrational and rotational energy of the molecule upon substitution of one isotope for another, has in the past proved capable of describing the displacements as accurately as they could be observed. With the discovery of an isotope of hydrogen of atomic weight 2, however, more rigorous tests of the theory have become possible. The ratio of the masses of the hydrogen isotopes is so much greater than that for isotopes of other elements that the resulting displacements in the hydride spectra are much larger. Consequently, the energy constants may be compared with greater accuracy, higher power terms may be studied and electronic energy displacements are not entirely negligible.

In the elementary theory the energy of the molecule is considered as made up of three parts—electronic energy, vibrational energy and rotational energy:

$$E = E_e + E_v + E_r.$$

The rotational and vibrational energy levels of a given molecule may generally be well represented by finding suitable values for constants in expressions of the form suggested by wave-mechanics. Thus the rotational levels in a given electronic and vibrational state are described by a polynomial in $J(J+1)$, where J is the rotational quantum number

$$E_r/hc = F(J) = B_v J(J+1) + D_v J^2(J+1)^2 + F_v J^3(J+1)^3 \ldots.$$

B_v, D_v, F_v, ... are constants for a given electronic and vibrational state but vary from one state to another. For the various vibrational states of a given electronic state they may be expressed as functions of v, the vibrational quantum number; for example:

$$B_v = B_e - \alpha(v+\tfrac{1}{2}) + \gamma(v+\tfrac{1}{2})^2 - \ldots,$$
$$D_v = D_e + \beta(v+\tfrac{1}{2}) \ldots.$$

Higher terms may be added to these expressions but are not as a rule necessary. The vibrational energy for the electronic state is described in a similar way by a series in $(v+\frac{1}{2})$,

$$E_v/hc = G(v) = \omega_e(v+\tfrac{1}{2}) - x_e\omega_e(v+\tfrac{1}{2})^2 + y_e\omega_e(v+\tfrac{1}{2})^3 \ldots,$$

where ω_e, $x_e\omega_e$, $y_e\omega_e$, ... are constants for the given electronic state. Observed levels correspond to $v = 0, 1, 2, \ldots$. B_e and ω_e are interpreted as values at an unrealizable vibrationless, rotationless equilibrium position of the molecule. B_e is connected with the corresponding moment of inertia of the molecule by the expression

$$B_e = h/8\pi^2 c I_e,$$

where h is Planck's constant, c is the velocity of light *in vacuo*. ω_e is taken as the frequency of vibrations of infinitesimally small amplitude about the equilibrium point. Ultimately these constants must depend on the constants of the nuclei and the number of electrons, but they are calculated in terms of an assumed internuclear distance r_e, the masses M and M' of the atoms, and the constants of a plausible law of force between the atoms. The other constants, which are of much smaller magnitude than B_e and ω_e, may be expressed as functions of these two. The masses M and M' enter the expressions for the energy in the form of μ, the reduced mass, where

$$\mu = MM'/(M+M').$$

In the elementary theory of the isotope effect, any dependence of r_e and the law of force upon the masses of the nuclei is neglected, and the constants determined from the spectrum of one isotope are corrected for the change in μ and compared with the values determined directly from the spectrum of the other isotope. This is most easily done by multiplying each constant by the appropriate power of ρ, where

$$\rho = (\mu/\mu')^{\frac{1}{2}}.$$

The reduced mass for the deuteride is written μ' and the other constants are distinguished in the same way.

If the complete energy expression[1] is written

$$E/hc = \sum_{l=0,1,\ldots} \sum_{j=0,1,\ldots} Y_{lj} (v+\tfrac{1}{2})^l [J(J+1)]^j,$$

then the correct power is given by

$$Y'_{lj}/Y_{lj} = \rho^{(l+2j)}.$$

Thus ω_e, B_e, α, ... are replaced by $\rho\omega_e$, $\rho^2 B_e$, $\rho^3\alpha$, ..., etc. The theory is tested by determining values of these constants from the isotopic spectra and comparing their ratios with these powers of ρ.

A method of treating the data derived from a particular spectrum in order to make an accurate comparison with the theory has been worked out by Jenkins and McKellar[2] in their study of the isotope effect for $B^{10}O$ and $B^{11}O$. Most of the work on the deuterides has followed the same general lines. Briefly the method is as follows. Each line of the spectrum corresponds to the transition of the molecule from a state

$$T' = T_e' + G'(v') + F'(J'),$$

to a new state

$$T'' = T_e'' + G''(v'') + F''(J'').$$

The lines which constitute a "band" of the spectrum arise from all possible changes from $F'(J')$ to $F''(J'')$ associated with a constant change in the electronic and vibrational energy. Thus, in the P and R branches of a band the lines $P(J)$ and $R(J)$ are given by the transitions $F'(J-1)' \rightarrow F''(J)''$ and $F'(J+1)' \rightarrow F''(J)''$ respectively. The differences between the wave-numbers of these lines lead to values of B_v and D_v for the band, since

$$\begin{aligned} R(J) - P(J) &= F'(J+1) - F'(J-1) \\ &= 4B_v'(J+\tfrac{1}{2}) + 8D_v'(J+\tfrac{1}{2})^3 + \ldots \end{aligned}$$

and $$R(J-1)-P(J+1)=F''(J+1)-F''(J-1)$$
$$=4B_v''(J+\tfrac{1}{2})+8D_v''(J+\tfrac{1}{2})^2+\ \ldots.$$

Values of the constants are chosen which most accurately represent the series of observed differences. Each set of values for B_v' and B_v'' is then used to calculate the constants of the second formula,

$$B_v=B_e-\alpha_e(v+\tfrac{1}{2})+\gamma_e(v+\tfrac{1}{2})^2\ \ldots.$$

The resulting values of B_e' and B_e'' determined from each of the two isotopic spectra then provide a test of the equation

$$B_e^i/B_e=\rho^2.$$

The D_e values provide a second but less accurate test.

The formulae for vibrational energy are tested in a similar way. Lines of the same branch and of the same J value are taken for successive bands. These correspond to different changes of the vibrational quantum number. If two bands with the same value of v' but differing by one unit in v'' are considered, then

$$R^{(v'v'')}(J)-R^{(v'v''+1)}(J)=P^{(v'v'')}(J)-P^{(v'v''+1)}(J)$$
$$=\Delta G''(v''+\tfrac{1}{2})-(B''_{v''}-B''_{v''+1})(J+\tfrac{1}{2})^2+\ \ldots$$
$$=\Delta G''(v''+\tfrac{1}{2})-\alpha''(J+\tfrac{1}{2})^2+\ \ldots.$$

These values are then used to give the constants in the formula

$$\Delta G(v+\tfrac{1}{2})=\omega_e-2x_e\omega_e(v+\tfrac{1}{2})+\ldots.$$

The values of ω_e thus derived from two isotopic spectra give a ratio ω_e^i/ω_e which may be compared with the value of ρ determined from the rotational structure and with the value calculated from the reduced masses.

Any remaining displacement between the lines of the isotopic spectra which is not a function of J and v is referred to as the electronic effect.

Already observations have been made on a large number of deuteride spectra. Jeppesen[3], Beutler and Mie[4], Beutler[5], Mie[6], Ashley[7], Dieke and Blue[8] have all made contributions to the analysis of the extremely complex spectra of the simple molecules H_2, HD, and D_2. The first three investigators have studied the bands of the extreme ultra-violet: the others have worked on the visible spectrum. Jeppesen identified twelve bands of HD corresponding to the $1s\sigma\,2p\pi\ {}^1\Pi_u \rightarrow 1s\sigma^2\ {}^1\Sigma^+_g$ system of H_2 in the spectrum obtained from a discharge through mixtures of hydrogen and deuterium. Mie, using a discharge through argon containing small amounts of hydrogen and deuterium, observed selective rotation doublet series of HD in the region λ 1650–λ 1000. The values of B and ω which he observed for HD agree well with those calculated from the isotope theory. The residual electronic effect which he found amounted to 15 ± 10 cm.$^{-1}$ Dieke and Blue have studied the visible spectrum, using a heavy-current discharge tube containing various mixtures of hydrogen and deuterium, and have assigned lines to H_2, HD or D_2 by studying their variation in intensity. They have studied in detail the isotope effect in the Fulcher bands, $1s\sigma\,3p\pi\ {}^3\Pi \rightarrow 1s\sigma\,2s\sigma\ {}^3\Sigma$, and conclude that although all of the constants vary with the proper power of ρ, yet there are outstanding deviations from the

values calculated using Bainbridge's values 1·007775 for the mass of the proton and 2·01363 for the mass of the deuton. For the separation of the origins of the systems, the electronic effect, they obtained the value 5 cm.$^{-1}$, which agrees with the result of Mie and with the later result of Jeppesen, 9 ± 5 cm.$^{-1}$ Other investigations include a study of the alternation of intensities in the lines of the alpha bands of D_2, by Lewis and Ashley[9], which led these authors to conclude that the nucleus of deuterium obeys the Einstein-Bose statistics; and the measurement of the ratio of the intensities by Murphy and Johnston[10] which enabled them to determine the spin of the deuterium nucleus as equal to 1.

The spectra of $Li^{(7)}H$ and $Li^{7}D$ have recently been investigated in very considerable detail by Crawford and Jorgensen[11]. The first part of their work deals with the analysis of 35 bands of the $^1\Sigma \rightarrow {}^1\Sigma$ system of $Li^{(7)}D$. The spectrum was obtained in absorption and is of the headless, many-lined type. It stretches from λ 3200 to λ 4300. The second part of the work consists of a reinvestigation of the corresponding system of $Li^{(7)}H$ and a preliminary comparison of the constants with those of $Li^{(7)}D$. Here also definite departures are found between the calculated and observed values of ρ. The authors find that the theoretical value of ρ, namely 0·750486, has to be increased by $0{\cdot}00014 \pm 0{\cdot}00003$ to give the observed values of ω_e'/ω_e, and by $0{\cdot}00020 \pm 0{\cdot}00006$ to give the observed ratio of B_e'/B_e for the final $^1\Sigma$ state. The data for the initial state present difficulties on account of the anomalous behaviour of this level which is a common feature of the upper levels of the hydrides of the alkali metals but is not observed in other diatomic spectra. There is no conclusive evidence of an electronic shift, their value being $0{\cdot}0 \pm 0{\cdot}5$ cm.$^{-1}$ The corresponding spectrum of NaD has been observed by Olsson[12] who finds, similarly, that the value of ρ calculated from the atomic weights has to be increased to give the ratio observed from B_e'/B_e. The calculated value is $\rho = 0{\cdot}72213$ and the observed value is $\rho = 0{\cdot}72236$.

In the next column of the Periodic Table, observations on BeD have been made by Koontz[13], and on CaD by Watson[14]. Koontz has photographed the $^1\Sigma \rightarrow {}^1\Sigma$ bands of BeD^+ from λ 2200–λ 3100, and also the $^2\Pi \rightarrow {}^2\Sigma$ bands of BeD at λ 4990. For the latter spectrum, photographed in the third order of a 21-foot concave grating, an accurate comparison between BeH and BeD has been made. From the atomic masses, $\rho^2 = 0{\cdot}55070$, while the ratio B_e'/B_e for the final state is 0·55110. The increase in ρ is here 0·00025. The rotational constants obtained for the final levels of the BeD^+ bands have also been compared with those found by Bengtsson for BeH^+. The resulting value of B_e'/B_e is 0·55073. An electronic shift of 0·8 cm.$^{-1}$ was observed. Watson has analysed the $B^2\Sigma \rightarrow X^2\Sigma$, orange system, of CaD, as well as the $C^2\Sigma \rightarrow X^2\Sigma$, ultra-violet system. He finds that the ratio B_e'/B_e is 0·51337 for the ground state, whereas the value from the reduced masses is 0·51276. The observed ρ is again greater, the increase in the value of ρ being 0·00043. The ultra-violet bands of MgD corresponding to the $B^2\Pi \rightarrow X^2\Sigma$ system of MgH have been investigated by Guntsch[15]. This system is interesting in affording a very striking example of predissociation. The $A^2\Pi \rightarrow X^2\Sigma$ bands of CdH and CdD have been observed and compared by Svensson[16].

Comparison of the AlD spectrum with that of AlH has been made by Holst and Hulthén[17] for the $^1\Pi \rightarrow {}^1\Sigma$ system. The constants for the lower level, $^1\Sigma$, yield the following ratios

$$\frac{B_e\,(\mathrm{AlD})}{B_e\,(\mathrm{AlH})} = 0{\cdot}51882,$$

$$\frac{\omega_e{}^2\,(\mathrm{AlD})}{\omega_e{}^2\,(\mathrm{AlH})} = 0{\cdot}51889,$$

whereas calculation from the atomic weights gives the values $0{\cdot}51848 \pm 0{\cdot}00002$.

Other deuteride spectra which have been observed include those of CD, ND, OD and DCl. Bands due to CD have been observed by McKellar and Bradley[18] both in the hollow cathode discharge through C_2D_2 and in the flame of burning C_2D_2 with excess of oxygen. Two bands appear at λ 4307 and λ 3865, corresponding to the $^2\Delta \rightarrow {}^2\Pi$ band of CH at λ 4315 and the $^2\Sigma \rightarrow {}^2\Pi$ band at λ 3871. The band of ND corresponding to the $^1\Pi \rightarrow {}^1\Delta$ band of NH at λ 3240 has been found by Dieke and Blue[19] at λ 3235. The Λ-doubling of the $^1\Pi$ levels is easily measurable for this system. As is expected from theory, the doublet separation is larger for NH, the ratio found being 3·511. The OD spectrum has been photographed by Johnston and Dawson[20], by Johnston[21] and by Shaw and Gibbs[22]. The Λ-doubling of the $^2\Pi$ levels is found to be smaller in the case of OD, being only 0·284 that of OH for equal values of K. The doubling of the $^2\Pi$ levels due to the electronic spin coupling has also been compared for these two molecules and is found to be greater for OD than for OH. The infra-red spectrum of DCl was measured by Hardy, Barker and Dennison[23], who found very satisfactory agreement between the calculated and observed positions of the absorption lines of the deuteride.

From these results, obtained from studies of a very varied selection of diatomic deuterides, it is apparent that although the elementary theory is capable of describing the observed effects to a high order of accuracy (to one or two parts in ten thousand) beyond this there is a definite departure. Observations from many spectra agree in giving a larger ratio between the constants than is to be expected from the theory. Many qualitative suggestions have been put forward to account for these discrepancies, but so far the work of calculating the magnitude of the corrections and comparing them with observations has not proceeded far enough to decide which are the important ones. Holst and Hulthén have suggested the inclusion, in the usual term for the moment of inertia, of a term for the moment of inertia of the electron cloud. Kronig[24] has pointed out that for many states there is to be expected a slight difference in the equilibrium distances for the hydride and deuteride, even for the non-rotating, non-vibrating molecules, when certain terms in the Hamiltonian which are usually neglected are taken into account. In the simple theory certain effects are neglected which have the result of adding small corrections to the coefficients of the energy formulae, and in general the small corrections require transformation by a different power of ρ on substitution of the deuton for the proton. Such interactions are the perturbation of one electronic level by another[25], the reaction of the nuclei to the precession of the electronic

angular momentum about the internuclear axis, and the interaction of rotation and vibration[26]. Dieke[27] has considered the last two effects in connexion with the H_2, HD and D_2 molecules, and has calculated that the reaction of the nuclei to the precession of the electronic angular momentum gives a term

$$[B_e-\alpha(V+\tfrac{1}{2})][L(L+1)-\Lambda^2],$$

which accounts very well for the large electronic shifts. Calculations of the effect of higher order approximations in the energy values for the rotating anharmonic oscillator have not yet been carried through in detail.

§ 3. ASTRONOMICAL SPECTROSCOPY

By A. Hunter, D.I.C., F.R.A.S.

Solar corona. The spectrum of the corona consists of several bright lines, probably some angstroms wide, superposed on a continuum. The presence of Fraunhofer absorption and of partial polarization in this continuum leaves little doubt that it is due, at least in part, to the reflection of ordinary sunlight from small particles extending to several solar diameters from the Sun's limb. Ever since the first satisfactory observations were made sixty-five years ago, speculation has been prevalent regarding the nature both of these particles and of the emitters responsible for the line spectrum. During the past year there have been published theoretical discussions, based upon material obtained at recent eclipses, which would appear to bring solutions to both these problems appreciably nearer.

Dealing first with the continuous spectrum: Grotrian[1] has been able, by a photometric study of the plates obtained at the 1923 eclipse, to show that this consists of a Fraunhofer spectrum superposed on a true continuum. The intensity of the Fraunhofer spectrum decreases slowly with increasing distance from the Sun, whilst the true continuum, though far more intense at the limb, loses brightness much more rapidly. It is this differential fading which led to the earlier division of the corona into an inner zone giving a truly continuous spectrum, and an outer zone giving a Fraunhofer spectrum. It is probable that both spectra originate in ordinary photospheric light, the difference between them lying in the different scattering processes suffered by this light in the coronal region itself. Whereas the Fraunhofer spectrum is probably due to ordinary "dust-reflection", i.e. scattering by particles large in comparison with the wave-length of light, the true continuum is to be regarded as due to scattering by free electrons whose thermal velocities are large enough to obliterate the absorption lines by Doppler broadening.

In an important paper[2] on the coronal lines, Goudsmit and Ta-You Wu point out that the great width of these lines is evidence that the emitting atom has an extremely short life in its excited state. This in turn suggests that the excitation may be of the type in which two electrons are simultaneously removed to outer orbits, leaving the atom in a state of higher energy than the ionization limit. From an approximate calculation of the term values which would be expected if the helium atom were excited in this way, it is predicted that several diffuse lines should lie in

the visible region. A theory of this type seems more satisfactory than the earlier explanation, which sought to identify "coronium" with some known element in a highly ionized state. Nevertheless, it has, of course, its disadvantages: for instance, it does not explain why ordinary helium lines do not appear in the coronal spectrum[3]. But in its support it should be mentioned that quite recently the green coronal line has been found in emission in the upper chromosphere[4], where, of course, helium lines are very prominent. Moreover, the five coronal lines which appeared in RS Ophiuchi in 1933 bore a strong resemblance in contour to the He II line at $\lambda 4686$. The next move may lie with the practical spectroscopist, who may have to face the problem of producing the coronal lines from helium in the laboratory. McLennan's production of the green auroral line in oxygen provides a precedent.

Fraunhofer spectrum and terrestrial absorption. Owing to the high intrinsic intensity of the solar spectrum in the infra-red, that spectrum has been among the first selected for systematic investigation by means of the newly-developed infra-red emulsions. First-fruits of the new technique include the discovery of phosphorus in the reversing layer[5], the confirmation of the presence of neutral sulphur there[6], and the assignment of many of the fainter infra-red lines in disc and spot spectra to iron[7]. In the far infra-red, where photographic recording is as yet impossible, such detailed results cannot, of course, be expected. Research in this region is of importance chiefly in telling us the constituents of the terrestrial atmosphere at heights at present unattainable by more tangible probes than sunlight. Recent investigations have shown bands at 6–7μ. due to H_2O; at $7{\cdot}4\mu$., 10μ., and 13μ. due to O_3, and at $13{\cdot}5\mu$. due to CO_2[8,9]. It is these last two bands which combine to set the present long-wave limit of transmission at $13{\cdot}5\mu$. Beyond this point the CO_2 band blends with the pure rotation spectrum of H_2O at about 17μ. to absorb all further radiation until at least 21μ.[9]

Unfortunately for terrestrial observers, study of the solar spectrum is seriously limited in the other direction, too, by the presence in the Earth's atmosphere of ozone and oxygen, both of which absorb strongly in the ultra-violet. The Hartley absorption band of ozone at $\lambda\lambda 3300$–2200 A. sets the ultra-violet limit of spectroscopic observation of solar radiation at $\lambda 2900$. Below $\lambda 2200$, oxygen absorption interferes with observation: a weak continuum extends beyond the Hertzberg bands near $\lambda 2500$, and there is almost complete absorption by the Schumann-Runge system below $\lambda 2000$. The most promising region for detection is therefore $\lambda\lambda 2200$–2000. A recent survey[10], however, shows that the amount of O_3 at ground level is too great to allow spectroscopic observation in this region, though at 16 km.—a height attainable by manned stratosphere balloons—this would be possible. Ozone unfortunately extends far too high for any hope to be entertained of avoiding absorption in the near ultra-violet[11]. There remains the possibility of detecting the spectrum by other means. Here success has been reported within the last few months[12], a monochromator having been used in conjunction with a sensitive photon counter to record the spectrum near $\lambda 2100$. It is very important that this work should be followed up, and the reality of the result placed, if possible, beyond criticism.

Limb effect. It has long been known that after the Doppler shifts caused by the solar rotation have been allowed for, the wave-lengths of Fraunhofer lines are, in general, greater at the Sun's limb than at the centre of the disc by amounts of the order of 0·005 A. The least unsatisfactory hypothesis yet put forward ascribes this difference to the effect of radially ascending currents in the Sun's atmosphere at the level of origin of the lines concerned. This mechanism would give lines from the centre of the disc a Doppler displacement to the violet relative to those from the limb, where the motion is across the line of sight. After allowing for our foreshortened view of the Sun, we may predict the way in which Fraunhofer wavelengths should vary across the disc. The observed variation, however, is very different from any such prediction, whatever the velocity of ascent assumed[13]. This fact led to the discrediting of the theory of radial currents until a recent suggestion by McCrea and Mitra[14] threw fresh light on the question. These authors point out that we can see deeper into the solar atmosphere at the centre of the disc than at the edges. Since the equation of continuity requires an ascending velocity which is inversely proportional to the density, this means that we see down to levels of lower velocity at the centre. The effective velocity of outward flow is therefore not the same all over the disc, and it is shown that for an idealized adiabatic atmosphere the theory may be partially reconciled with observation. A more detailed investigation may remove the remaining discrepancy.

Spectra of the night sky and aurora. Advances in the design of wide-aperture spectrographs have been recently utilized for extensive studies of the spectra both of the aurora and of the night sky. It is now possible to photograph the night-sky spectrum in six hours[15], and the auroral spectrum in an hour or less[16]. As is so often the case, attention which was first drawn to the similarity between the two spectra is now focussed on the differences they exhibit. The main features of the two spectra are the same: on a weak continuous background are superposed three bright lines (the famous forbidden line of neutral oxygen at $\lambda 5577$, and two red ones at $\lambda\lambda 6303, 6368$), and a number of bands, the majority of which can be attributed to the molecules N_2 N_2^+, O_2, and H_2O. Of the nitrogen bands the auroral spectrum contains the N_2 first positive group in the red and infra-red, the N_2 second positive system in the violet, and the N_2^+ first negative system in the blue, whereas the two last-named groups are very faint or even non-existent in the spectrum of the night sky. The first is, however, very much stronger, relative to the green line, in the night sky than in the aurora. It is of interest, too, that in the night-sky spectrum this first positive system exhibits the selective strengthening of certain bands which has hitherto been associated only with the afterglow of active nitrogen[17]. In the case of the afterglow, it is known that the abnormal population of the vibrational levels concerned is caused by collisions of the molecules with metastable nitrogen atoms produced in the exciting discharge. Whether the same agency produces the effect in the Earth's upper atmosphere is not certain, and elucidation of this point will no doubt add considerably to our knowledge of excitation mechanisms in the stratosphere. It seems probable that detailed study of the intensity differences between auroral and night-sky spectra[18], and of the nightly[19] and yearly[20] variations in

group intensities within each spectrum, will prove of equal importance in this respect.

Perhaps the most interesting discovery in this field during the last year has been the identification[21] of several weak bands appearing in both the auroral and night-sky spectra with some found independently by Vegard in the luminescence of solid nitrogen[22] and by Kaplan in a new nitrogen afterglow[23]. Analysis of the bands shows that they are the long-sought $A \rightarrow X$ intercombination bands which fix the relative positions of the triplet and singlet levels of the neutral nitrogen molecule. The presence in the upper atmosphere of molecules in the metastable triplet state A has, of course, long been known, since this is the final level of the first positive group, one of the most prominent features of both auroral and night-sky spectra. Such metastable molecules have relatively long lives which, it has hitherto been assumed, must be ended by degrading collisions with other particles, since no transitions to the singlet ground state X involving radiation are allowed by the selection rule. We may now suppose, however, that the mean free path of the excited molecules is so large at the height at which luminescence occurs, that they have a fair chance of returning to the ground state by radiating the "forbidden" bands before they suffer collisions.

Interstellar matter. As was mentioned in this report last year, there is now overwhelming evidence that in addition to the condensations we know as stars and nebulae, diffuse clouds of atoms are distributed, probably fairly uniformly, throughout galactic space. The presence of these clouds is shown by the so-called "detached" lines, which show no Doppler shifts, appearing in absorption in the spectra of very distant stars. Another type of absorption in interstellar space is that giving rise to "colour excess" in distant stars: amongst the members of a given spectrum class, the most distant stars appear reddest. It is obviously of interest to find how the light-scattering material responsible for this phenomenon is distributed in space. A solution to this problem is suggested by the recent work of Williams[24], who has compared the colour excesses of sixty-seven B- and A-type stars with the intensities of their interstellar K-lines (due to ionized calcium). He finds a statistical correlation between these quantities, but concludes that the scattering material and the interstellar calcium are not coextensive. The Be stars, for example, are all redder than the intensities of their calcium lines would indicate; such bright-line stars are probably surrounded by a local concentration of scattering matter ejected by their rapid rotation. Examination of sixty-two stars of types O and B for colour excess leads Thorndike[25], also, to attribute the differences he finds to scattering by irregularly distributed clouds.

The need for caution in interpreting results which assume a linear relation between distance and interstellar line intensity is emphasized by the work of several investigators. By means of a statistical comparison between stellar distances given by such a relation and those found from the galactic rotation effect, Struve[26] is able to show the existence of local condensations and rarefactions in the cloud of interstellar calcium. Beals[27], too, finds differences of interstellar intensities amongst stars in the same cluster, which can obviously not be due to varying distances. But

in this case it is equally evident that the discrepancies cannot be put down to local variations in density of the calcium cloud. A blending with stellar calcium lines is suggested as the most reasonable explanation.

Since last year's report was published, many more instances[28] of interstellar absorption of the light of late-type stars have been added to the solitary first example recorded then. In this way, one of the chief objections to the theory of interstellar clouds—that only early-type stars show detached lines in their spectra—is being gradually removed. It seems probable that most of the failures to detect the effect in late-type stars are due to the superposition of sodium and calcium lines of stellar origin on the absorption lines due to the intervening clouds. If this is so, it would seem necessary merely to select stars in whose spectra separation of the two kinds of lines is likely: such as very distant stars with high radial velocities, or very distant eclipsing binaries in which the stellar lines are removed at certain phases of the orbit.

Another outstanding objection to the theory of diffuse atomic clouds has been dealt with by Eddington. His original calculations appeared to show that for no reasonable interstellar abundance of sodium would the intensities of the D-lines approach those of the H- and K-lines of ionized calcium. In a more recent paper[29], however, he examines in greater detail the relation between the intensity of an interstellar line and the density of the absorbing atoms responsible for its production. It now appears that for only very low and very high densities does the intensity rise appreciably with density. This rise is due to increasing absorption, at the centre of the line for low densities, and on the edges of the line for high densities. The middle range of densities, in which central absorption is complete before wing absorption sets in, happens to include at its extremes the densities of neutral Na and ionized Ca which, it is suggested, actually occur in space. Hence although these densities differ so much, little difference should be expected in the intensities of the corresponding lines. This conclusion is supported by the measurements of Beals[27].

Novae. The spectrum of the recurrent nova RS Ophiuchi, whose 1933 outburst was described in this report last year, has now reverted to its normal appearance, with strong lines of ionized helium and iron superposed on a strong continuum. The five coronal lines which were such an interesting feature of the spectrum in its earlier stages have now completely disappeared.

December 1934 was marked by the appearance of a naked-eye nova in Hercules. The newcomer reached its maximum brightness on December 22, when its magnitude was 1·4[30]. Its development followed the normal course fairly closely, the hydrogen lines H_α to H_ϵ showing the absorption on their violet edges which is characteristic of novae[31][32]. The current interpretation of this absorption is that it occurs in a layer of relatively cool gas receding radially from the star as a result of the initial outburst. By the end of January 1935, the velocity of this layer, as shown by the Doppler shift of the hydrogen absorption lines, had increased to 900 km./sec.[32]. Other lines, however, do not show this type of absorption; two bright maxima, almost equally displaced to red and violet, suggest, rather than an expanding shell of gas, streams or jets of equal velocities in opposite directions[33].

Perhaps the feature of most significance to the student of stellar evolution will prove to be the development in mid-February of three forbidden lines of neutral oxygen[33], including the auroral green line, which correspond exactly in term relationships to the nebular forbidden lines of O III. This is the first time that the auroral line has appeared in a stellar source. The observation is likely to be even more valuable than that of the coronal lines in Nova Ophiuchi in 1933, since the excitation potentials of the lines are known, and the physical conditions necessary for their emission can actually be reproduced in the laboratory.

§ 4. INFRA-RED BAND SPECTRA OF ORGANIC MOLECULES

By LOTTE KELLNER, Dr.Phil.

The first investigations of the infra-red spectra of organic molecules were carried out in 1892 by H. Julius[1] who found that molecules containing the group CH_3 or CH_2 showed an intense absorption band at 3μ. ($1\mu. = 10^{-4}$ cm. = 10,000 A.). Thirteen years later W. W. Coblentz[2] published the results of his extensive experiments on the infra-red absorption spectra of a great number of organic compounds. He varied systematically the constituents in the molecules and studied their influence on the spectra. The result of these studies has been the foundation of our knowledge of this subject. Since then a considerable amount of work has been done, chiefly by Lecomte and his collaborators in Paris and by Ellis and others in America[3,4].

A few words should first be said about the technique employed in infra-red spectroscopy. As we are dealing here with long wave-lengths (from 1μ. up to 20μ.), photography is impossible, and a photometric method has to be adopted. Usually the resolution of a prism-spectrometer is sufficient. The intensity of the radiation is measured by means of a thermopile—preferably a vacuum-thermopile—in connexion with a high sensitivity galvanometer. The material of which the prism is made varies according to the region of the spectrum which is to be investigated. From 1 to $2{\cdot}5\mu$. a glass prism is used, from $2{\cdot}5$ to $3{\cdot}5\mu$. quartz, from $3{\cdot}5$ to 9μ. fluorite and from 9μ. to 20μ. rock salt, to ensure the greatest possible resolution. In order to obtain the rotational structure of the bands, it is necessary to employ a grating. In this case the sensitivity of a thermopile will not be sufficient; either an amplification of the galvanometer current must be provided or a receiving instrument of higher sensitivity (radiometer, radiomicrometer) has to be used. As it is impossible to observe the emission spectra of organic molecules, owing to their weak intensity, we always deal with absorption spectra. As light-sources, tungsten filament lamps for the glass region and Nernst glowers for longer wave-lengths, are generally used. The absorption of the substances is very strong: cell thicknesses of 0·1 to 0·01 mm. are sufficient.

As the structure of organic molecules is highly complicated, it is to be expected that their spectra will also be extremely complicated. On the other hand, it simplifies the matter that the infra-red molecular spectra are only connected with

changes of the vibrational and rotational energy of the molecule in which the electronic structure is not changed. As long as we confine ourselves to the study of the unresolved bands we are dealing only with the changes of the nuclear arrangement in the molecule. Nevertheless, if we consider a molecule consisting of n atoms ($n > 2$), it will have $3n-6$ inner degrees of freedom, and consequently $3n-6$ independent vibrations. Not all of these vibrations will be active in the infra-red spectrum; only those frequencies have a non-vanishing intensity which are caused by a change of the dipole moment. Besides, two or more vibrations may have the same frequency, which fact would further diminish the number of the actually occurring bands. But as the same groups of atoms appear in many organic compounds (for example the groups CH_3, CH_2, OH, etc.), it is quite feasible that molecules of similar structure will have spectra which show corresponding similarities. As a matter of fact, a number of bands appear in the same positions for many organic molecules. Table 1 gives the wave-numbers for the absorption bands of several hydro-carbon compounds C_nH_{2n+2}. Ethane, propane and butane have been studied in the gaseous state, the other three as liquids.

Table 1 shows that certain groups of bands occur in the spectrum of each member

Table 1

Ethane (2) cm.$^{-1}$	Propane (5) cm.$^{-1}$	Butane (2) cm.$^{-1}$	Hexane (2) cm.$^{-1}$	Octane (2) cm.$^{-1}$	Dodecane (2) cm.$^{-1}$
—	735	—	724	724	722
750	750	752	757	757	741
—	—	—	—	—	763
800	—	—	806	813	809
820	—	—	—	—	—
837	—	833	—	—	—
—	865	893	885	885	885
—	—	—	—	—	944
—	930	—	967	959	970
—	—	1003	1010	—	—
—	—	1042	1062	1073	1080
—	—	1124	1134	—	—
—	1170	—	1163	1160	1160
—	—	1205	—	—	1220
—	—	—	1282	1282	1285
1379 [6]	—	—	1360	1364	1364
—	—	—	—	—	1424
1460	1450	1451	1458	1456	1458
—	—	—	—	—	1528
—	—	—	—	1614	1620
—	1730	1754	1721	1718	1730
—	—	2222	2381	2500	2500
—	2900	—	—	—	—
—	—	2924	2924	2924	2924
2950	2970	—	—	—	—
—	3190	—	—	—	—
—	3350	—	—	—	—
—	3480	—	—	—	—
—	3700	—	—	—	—
—	3700	—	—	—	—
—	3820	—	—	—	—
—	4080	—	—	—	—
4237	4240	4167	4132	4202	4219

of the homologous series. There is, as would be expected, a slight shift of wave-number in going from the gaseous to the liquid state. Some of these bands, common to all substances, appear in almost identical positions; others show a tendency to shift to higher frequencies with the heavier molecules. A computation of the mean values of the frequencies common to all molecules C_nH_{2n+2}, which have been investigated by Coblentz[2] and Ellis[7], gives the following results:

Bands occurring in all molecules C_nH_{2n+2}.

cm.$^{-1}$	cm.$^{-1}$	cm.$^{-1}$	cm.$^{-1}$	cm.$^{-1}$	cm.$^{-1}$	cm.$^{-1}$
724*	1075	1360*	1721	4292	5714	8403
810	1159	1459*	2915*	4955	5882	9756
885	1285	1605	4202	5376	7194	10869
964						

The strongest bands have been marked by an asterisk. As these molecules all consist of a chain of CH_2 groups which end in a CH_3 group on either side of the chain, the attempt was made to identify some of the observed frequencies with the inner oscillations of the groups or with the oscillations of the groups against each other. As a substitution for the hydrogen atoms in the CH_2 groups of one or more halogen atoms shows that the bands at 2915 and 1459 cm.$^{-1}$ decrease in intensity and eventually vanish completely, it is obvious that these bands are caused by vibrations of the hydrogen atoms in the CH_2 group, but as benzene has similar bands at 3077 and 1481 cm.$^{-1}$, it follows that the presence of one H-atom is sufficient to produce them. Formerly, the band at 1459 cm.$^{-1}$ was considered to be the fundamental and 2915 cm.$^{-1}$ the first overtone of the CH-valence vibration, i.e. a vibration of the hydrogen atom against the carbon atom along the valence bond. However, as the two bands have almost the same intensity, this interpretation cannot be correct. Therefore Ellis[8] assumed the band at 2915 cm.$^{-1}$ to be the fundamental. Correspondingly, one looks for a vibration of the C—C bonds which ought to appear in the region between 800 and 1100 cm.$^{-1}$, but it is still impossible to assign any of the bands in this region to a C—C valence frequency.* In analogy with the spectrum of methane, the frequencies at 1360 and 1459 cm.$^{-1}$ can be associated with movements of the hydrogen atoms which cause an alteration of the valence angle of the CH bond (bending frequencies).

As mentioned above, benzene and its derivatives have a spectrum which shows the same structure as that of the paraffins. The bands of benzene have the following positions: 777, 804, 819, 971, 1023, 1153, 1282, 1380, 1481, 1613, 1820, 2041, 2273, 3077, 3703, 4016, 4587, 5952, 6993, 8772, 9804 cm.$^{-1}$ The two bands connected with the CH group at 1481 and 3077 cm.$^{-1}$ appear at shorter wave-lengths than the corresponding bands of the aliphatic substances. That is probably due to the fact that the carbon atoms are bound by alternating double bonds in the benzene molecule.

Hitherto we have been considering compounds consisting of hydrogen and carbon atoms. Exhaustive investigations have been made on the infra-red spectra of derivatives of these compounds. The best known among them are the halogen

* Strictly speaking, there will be more than one C—C valence vibration for chains with more than two carbon atoms.

derivatives, the alcohols, aldehydes and ketones[3,4,10]. Experiments on these substances gave the following results:

The bands characteristic of the hydro-carbon compounds appeared in the spectra of the derivatives in slightly different positions. In addition, new bands occurred which were attributed to the inner vibrations of the substituted group or the vibrations of the substituted group against the binding carbon atom. For example, it was found that the substitution of a halogen atom for a hydrogen atom resulted in a band at 650 cm^{-1} for chlorine, 560 cm^{-1} for bromine and 500 cm^{-1} for iodine. These frequencies were assigned to the oscillation of the halogen atoms along the valence bond between the carbon and halogen atoms. Similarly, we find a very strong band at 1712 cm^{-1} for the fatty acids and at 1695 cm^{-1} for the aldehydes and ketones. We may assume accordingly that this band is caused by a vibration of the carbon and oxygen atoms along the C=O valence bond. On the other hand, the alcohols show very strong absorption at 3295 cm^{-1}. As water possesses an intense band in the same region, the corresponding frequency of the alcohols is generally attributed to the valence vibration of the OH-group. But it is remarkable that no frequency exists characteristic of the oscillation of the OH-group as a whole against the chain.

The most puzzling feature of the infra-red spectra of organic molecules is the fact that the constituents of the molecules are oscillating independently of the movements of the rest. We should expect that the substitution of a new group would alter the arrangement of the whole spectrum. Actually we find that the spectrum of the molecule is only very slightly changed by displacing one or more constituents by another atom or a group of atoms. In addition a number of frequencies appear which are characteristic of the substituted group. The first attempt to give a theoretical explanation of the general appearance of these spectra was made by Bartholomé and Teller in 1932. They treated the problem of computing the vibrations of a chain of CH_2-groups by classical methods. Three assumptions had to be made to simplify the calculations: (1) Harmonic forces act between the groups. (2) The groups CH_3, CH_2, OH etc. may be treated as mass points having the masses of the groups. This means that the inner degrees of freedom are neglected; the vibrations connected with them—in the region of 3000 cm^{-1}—vanish from the spectrum. (3) The potential energy is a function only of the distance between those atoms which are connected by a valence bond: the forces which maintain the angles between the valence bonds are not taken into account. The calculations for a chain of n mass points whose links form a tetrahedral angle with one another, furnish the result that the vibrations of the chain (valence frequencies) lie between an upper and a lower limit. These limits are computed to be 836 and 1185 cm^{-1}. A chain consisting of n C—C valence bonds has n valence frequencies, but as they lie within the same limits, independently of the length of the chain, they tend, when the length of the molecule grows, to merge into a continuous background, from which only the strongest bands will stand out. If a different mass point, bound by a different force to the neighbouring point of the chain, is now added, the chain will be excited to new oscillations in addition to those already existing, and there will also be vibrations of the added mass point relatively to the chain. If the former oscillations have

frequencies greater than the upper limit or smaller than the lower limit of the original chain frequencies, their amplitudes will decrease exponentially; in other words, these oscillations will die out, leaving the original vibrations of the chain, supplemented by vibrations of the added mass point against the chain. Computation shows that this should occur for the atoms Cl, Br, I and H, and this is confirmed by experiment. The groups, OH and NH_2, on the other hand, should not excite the chain to vibrations lying outside the above-mentioned limits; and in these cases no vibrations characteristic of these groups should occur relatively to the chain. This also agrees with experiment.

The interpretation of the infra-red spectra of hydro-carbon compounds would be facilitated by the knowledge of the rotational structure of the vibration-rotation bands, for the type of the fine structure depends on the type of vibration on which the rotational structure is superposed[11]. But owing to the high moment of inertia of the molecules with which we are concerned, only a few, relatively simple molecules have been successfully investigated so far, namely methane, ethylene, acetylene and formaldehyde[6, 12]. Levin and Meyer[6] attempted to resolve the fine structure of ethane, but did not completely succeed; obviously the resolving power of their spectrometer was not sufficient. It has been possible to interpret the spectra of these molecules and to assign a definite mode of vibration to the fundamentals. But the more complicated molecules still defy a complete analysis of their spectra, in spite of the great regularities in the distribution of the bands. As we are not able to decide which are the fundamental bands, we are equally unable to determine their overtones and combination frequencies. The CH- and OH-frequencies are exceptions. Though only the infra-red spectra have been considered here, it should be pointed out that an interpretation of the data is impossible without the additional knowledge of the Raman spectra.

The discovery of the heavy hydrogen isotope will very likely be a great help in the task of pushing forward the analysis of the molecular spectra of organic compounds. The introduction of a diplogen-atom into the molecule will result in an isotopic shift of the bands. The amount of the isotopic shift will be an indication which groups of the molecule are participating in the oscillation under consideration. Further progress may be expected from an extension of our knowledge into shorter wave-lengths. It is feasible, on the other hand, that a more complete theoretical treatment of complicated molecules will be forthcoming which will provide the experimentalist with greater possibilities of sorting out the different types of vibration from the confusing heap of data.

§ 5. THE HYPERFINE STRUCTURE OF SPECTRUM LINES AND NUCLEAR MOMENTS

By E. Gwynne Jones, Ph.D., A.Inst.P.

An unusually large number of nuclear moments (I) were determined, in 1934, by the analysis of hyperfine structures. At the present time, nuclear moments are known for some 60 nuclei and for a further 35 nuclei of even mass it can be said

that the nuclear magnetic moments are comparatively very small (possibly zero), so that it is not possible to assign nuclear spins on account of the absence of observable hyperfine structures. Following the usual convention, the spins of such nuclei are set equal to zero in Table 1. In some of these cases a study of the alternating intensities in band spectra has shown that the nuclear spins are actually zero.

Schüler and his collaborators have found it possible[7] to excite the spectra of many rare elements, in a modified form of the well-known Schülerlamp, with sufficient intensity to examine the line structures by means of Fabry-Pérot étalons; nuclear spins were thus assigned to protactinium, europium, holmium, lutecium, terbium, thulium and scandium. Protactinium is of especial interest, since it is the only radioactive element in which hyperfine structure has been observed and, in addition, only a few milligrams of the specimen were available for the whole of the investigation. Other workers[a-9] have determined mechanical moments (I) for cobalt, columbium, lanthanum, tin and xenon and nuclear magnetic moments (μ) for cobalt, columbium, fluorine, lanthanum, potassium and tin.

A notable advance in the technique of hyperfine structure observations was made by Jackson and Kuhn[9], who produced the potassium resonance lines by absorption in an atomic beam and discovered a hyperfine structure which had been hitherto masked by the Doppler breadth of the lines. By making observations at right angles to the direction of the atomic beam, lines with a Doppler width corresponding to a temperature of 4° Abs. were obtained.

A very satisfactory feature of the year's progress is the determination of I-values by several completely independent methods. For example, the nuclear spin of sodium has now been fixed by four methods—the magnetic resolution of atomic beams (Stern-Gerlach experiment), hyperfine structure, the polarization of resonance radiation and band spectra—all of which lead to the same result, $I=\frac{3}{2}$. Accordingly, an account of the newest methods will be given as part of the present *Report*.

Perturbations of hyperfine structures. In continuation of the general introduction given in the previous *Report*, some account must be given of the perturbations that occur in the hyperfine structures of the spectra Hg I, Pb II, Al II and Bi IV. The subject was, indeed, briefly mentioned (p. 116) but its importance warrants a discussion here.

The hyperfine structure levels of an atomic energy state are ultimately distinguished by the quantum number, F, which is a measure of the total moment of momentum of the whole atom, in units $h/2\pi$. As is to be expected on theoretical grounds, neighbouring levels having the same F-value may engage in an interaction (if they are not too far apart) whereby the separation of the two levels is modified. The process may be described by stating that levels of the same F-value, but belonging to different terms, tend to repel each other if sufficiently close.

From the point of view of hyperfine structure modifications, two different types of perturbations are to be distinguished according to the similarity, or otherwise, of the J-values characterizing the interacting states.

Case (i). When the J-values of the two terms are the same, the interaction be-

Table 1. Nuclear mechanical and magnetic moments

Atomic number Z	Symbol	Mass numbers	Nuclear spin I	Nuclear magnetic moment μ
1	H	{1	$\frac{1}{2}$	3·25}
		2	1	0·8 }
2	He	4	0	0
3	Li	{6	0	0 }
		7	$\frac{3}{2}$	3·29}
6	C	12	0	0
7	N	14	1	—
8	O	16	0	0
9	F	19	$\frac{1}{2}$	≈3
11	Na	23	$\frac{3}{2}$	2·0
13	Al	27	$\frac{5}{2}$	2·1
15	P	31	$\frac{1}{2}$	—
16	S	32	0	0
17	Cl	35	$\frac{3}{2}$	—
19	K	{39	$\frac{3}{2}$	0·39}
		41	$\frac{3}{2}$	0·16–0·3 }
21	Sc	45	$\frac{7}{2}$	—
23	V	51	$\frac{7}{2}$	—
25	Mn	55	$\frac{5}{2}$	—
27	Co	59	$\frac{7}{2}$	≈2
29	Cu	63, 65	$\frac{3}{2}$	2·5
30	Zn	{66–70	0	0 }
		67	$\frac{5}{2}$	— }
31	Ga	{69	$\frac{3}{2}$	2·01}
		71	$\frac{3}{2}$	2·55}
33	As	75	$\frac{3}{2}$	0·9
35	Br	79, 81	$\frac{3}{2}$	—
36	Kr	{78–86	0	0 }
		83	$\frac{9}{2}$ or $\frac{7}{2}$	— }
37	Rb	{85	$\frac{5}{2}$	1·3}
		87	$\frac{3}{2}$	2·7}
38	Sr	{86, 88	0	0 }
		87	$\frac{9}{2}$	1·0}
39	Yt	89	(?)	—
41	Cb (Nb)	93	$\frac{9}{2}$	3·7
47	Ag	107, 109	$\frac{1}{2}$	< 0·25
48	Cd	{110–116	0	0 }
		113, 115	$\frac{1}{2}$	−0·6}
49	In	115	$\frac{9}{2}$	5·4
50	Sn	{116–124	0	0 }
		117, 119	$\frac{1}{2}$	0·9}
51	Sb	{121	$\frac{5}{2}$	4·0}
		123	$\frac{7}{2}$	3·1}
53	I	127	$\frac{5}{2}$	—
54	Xe	{124–136	0	0 }
		129	$\frac{1}{2}$	— }
		131	$\frac{3}{2}$	— }
55	Cs	133	$\frac{7}{2}$	2·6
56	Ba	{136, 138	0	0 }
		137	$\frac{3}{2}$ (?)	0·9}
57	La	139	$\frac{7}{2}$	2·7
59	Pr	141	$\frac{5}{2}$	—
63	Eu	151, 153	$\frac{5}{2}$	—
65	Tb	159	$\frac{3}{2}$	—
67	Ho	165	$\frac{7}{2}$	—
69	Tu	169	$\frac{1}{2}$	—
71	Lu	175	$\frac{7}{2}$	—
73	Ta	181	$\frac{7}{2}$	—
74	W	{182–186	0	0 }
		183	$\frac{1}{2}$	— }
75	Re	187, 189	$\frac{5}{2}$	—
78	Pt	{194, 196	0	0 }
		195	$\frac{1}{2}$	— }
79	Au	197	$\frac{3}{2}$	1·8
80	Hg	{198–204	0	0 }
		199	$\frac{1}{2}$	0·55}
		201	$\frac{3}{2}$	−0·62}
81	Tl	203, 205	$\frac{1}{2}$	2
82	Pb	{204–208	0	0 }
		207	$\frac{1}{2}$	0·60}
83	Bi	209	$\frac{9}{2}$	0·89
91	Pa	231	$\frac{3}{2}$	—

tween their hyperfine levels is large and the effect is measurable over large distances. The effect is similar to, and occurs with, the perturbation of term series discovered by Shenstone and Russell[10] in 1932. When the process is summed over all the hyperfine levels that comprise a set of atomic states, it is seen that the result of the perturbation is a mutual sharing of properties between the interacting terms. If one term arises from an electron configuration involving penetrating electrons, the perturbation will increase the hyperfine separations of the second term and, generally, cause that term to behave as if it also arose from a penetrating electron. An example of this effect is found in the $^1P_1$-series in Hg I. Here most of the members show small hyperfine structures and no isotope shifts, but the third number, $8\ ^1P_1$, appears with

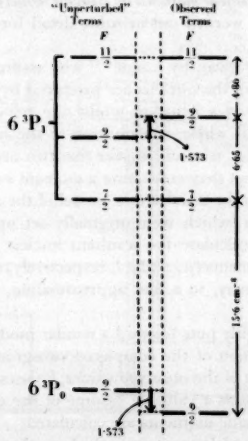


Figure 1. Perturbation of hyperfine structure levels in Bi IV.

a greatly widened structure (twice that of the lowest term of the series) together with an isotope shift. This apparent anomaly is caused by the proximity of a $^1P_1$-term of another configuration ($5d^9.6s^2.6p$) which involves two penetrating s-electrons—in fact just the type of term that is expected to show large isotope shifts and hyperfine structure.

Case (ii). The other class of perturbation (in which the J-values of the interacting terms are different) occurs very rarely and then only when the separation of the terms is of the same order as the hyperfine structure of one of them. Since the J-values differ, the range of F-values, from $(I+J)$ to $(I\sim J)$, is not the same in both terms, so that all the hyperfine sub-levels do not take part in the interaction. Hence relative displacements of the individual hyperfine levels are brought about and, as a result, the interval rule no longer holds. In the example shown, figure 1, we have two

of the terms, $6\,^3P_0$ and $6\,^3P_1$, arising from the configuration $6s6p$ of Bi IV. The $6\,^3P_0$ has only one hyperfine level ($F=\frac{9}{2}$), whilst $6\,^3P_1$ has three ($F=\frac{11}{2}$, $\frac{9}{2}$ and $\frac{7}{2}$). Now the separation of the two levels for which $F=\frac{9}{2}$ is of the same order as the overall hyperfine separation of $6\,^3P_1$, and so the perturbation between these levels produces a very large deviation from the interval rule in $6\,^3P_1$. Since the (unaltered) total separation of $6\,^3P_1$ is known, an application of the interval rule gives the "unperturbed" position of the $F=\frac{9}{2}$ level, which is found to have been displaced by 1·573 cm.[1] Had the perturbation not been recognized, in this case the F-values would have been incorrectly assigned, since the larger interval is expected to be between the levels with $F=\frac{11}{2}$ and $\frac{9}{2}$.

The representation of nuclear moments by a vector model of the nucleus. Two models have been proposed and worked out in some detail for all nuclei that have been investigated.

The simpler model is given by Landé[11] who assumes that all the mechanical and magnetic properties of the nucleus are produced by the orbital motion of one particle (either a proton or a neutron) whilst the remaining constituents of the nucleus form closed shells whose contribution to the nuclear moment is usually zero. In some special cases, it would appear that two neutrons combine with their spin vectors parallel, so that they contribute a moment $s=1$ to the resultant. Since this model is analogous to the usual vector model of the atom, it is possible to take over the Landé formulae (which were originally set up to deal with the orbital motion of electrons), to calculate the resultant nuclear magnetic moment (μ) in terms of the magnetic moments (μ_P and μ_N, respectively) of the proton and neutron. The observed moments may, to a first approximation, be represented by taking $\mu_P=2$ and $\mu_N=-0{\cdot}6$.

Schüler[12] independently puts forward a similar model, but adds a third vector to allow for the contribution of the completed sub-groups. In order to obtain a satisfactory representation of the observed values, he uses $\mu_P=2{\cdot}5$ and $\mu_N=-1{\cdot}65$. Schüler's paper also contains a valuable résumé of the existing experimental data from which nuclear magnetic moments are calculated.

The success of these models encourages the hope that it may soon be possible to formulate an accurate vector model, which will account completely for all the data on nuclear moments.

Independent methods of determining nuclear moments: (*a*) *The polarization of resonance radiation.* During the year 1934, three papers[13,14,15] have appeared on measurements of the resonance radiation in sodium and caesium. These, together with the earlier measurements on mercury and cadmium, show that excellent results are obtained when theory is corrected to allow for the effect of nuclear spin on the atomic states that give rise to the lines. In fact, the subject has now developed into an accurate study of the Zeeman effect of hyperfine structures in weak fields. A great simplification is made in the instrumental requirements since it is not necessary to separate the individual Zeeman components, and so separations that are smaller than the line widths may be deduced. Some idea of the importance of the new corrections to theory may be gathered from the fact that the degree of polarization

in the resonance lines of sodium (when excited by a beam of plane polarized sodium light) is 16 per cent, whilst on the older theory it should be 50 per cent. Taking the nuclear spin as $\frac{3}{2}$, the revised theory requires 15 per cent.

The experiment consists in passing plane polarized light into (say) sodium vapour and in measuring the amount of polarization in the resonance light emitted at right angles to the incident beam. The degree of polarization is modified when the resonance flask is placed in a magnetic field; the magnitude of the change depends upon the relative orientations of the plane of polarization in the incident light, the lines of force and the direction of observation.

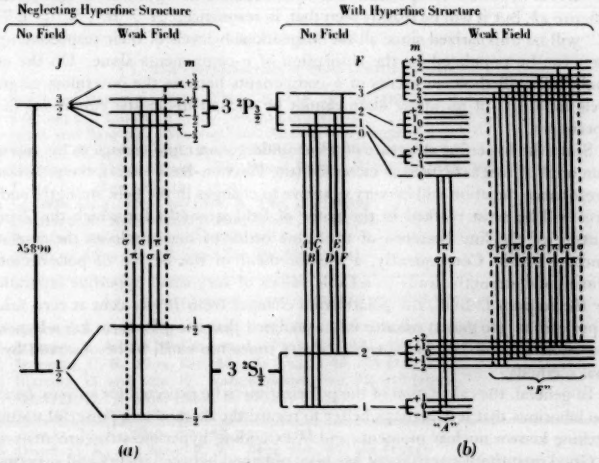


Figure 2. The polarization of Na I λ5890 produced by resonance in a weak magnetic field.

The polarization of the resonance radiation is best explained by a consideration of the process in a magnetic field, in the absence of which the system may be regarded as degenerate. Suppose the lines of force to be parallel to the electric vector in the exciting radiation, so that only the π-components of the Zeeman pattern can be absorbed. Since these represent the transitions for which the magnetic quantum number, m, does not change, it may happen that all the magnetic sub-levels* of the upper state are not excited by absorption from the ground state. Figure 2*a* shows how this can happen in the sodium D_2-line, λ5890. Since the two Zeeman components of the $3\,^2S_{\frac{1}{2}}$-term are so close together that absorption takes place from both and

* The "magnetic" sub-levels are those that arise in a magnetic field; transitions between them produce the Zeeman pattern (figure 2*a*). The π-components are polarized parallel to, the σ-components perpendicular to the field, when the direction of observation is perpendicular to the field.

only the π-components can be absorbed from the incident light, the sub-levels with $m = \pm\frac{3}{2}$ in the upper state will not be excited at all. Hence, in resonance, the outermost σ-components (shown by broken lines) will not be present in the line pattern, which will thus show an excess of π-components and appear partially polarized. The degree of polarization can thus be calculated if the relative intensities of the Zeeman components are known. When the hyperfine structure of the line is taken into account, a similar preponderance of π-components is found in the resonance emission, but the proportion is now different owing to the greater number of possible transitions. On account of the great complexity of the Zeeman effects of the various hyperfine components in a weak field, it is not possible to draw them all in figure 2*b*, but it will be readily seen that, in resonance, "*A*", "*B*", "*C*", "*D*" and "*E*" will be unpolarized since all the magnetic sub-levels of their respective upper states can be populated by the absorption of π-components alone. On the other hand, "*F*" will show an excess of π-components because the outermost magnetic levels ($m = \pm 3$) of its upper state cannot be excited under the conditions of the experiment.

Since the hyperfine structure of a line undergoes a rapid change as the intensity of an applied magnetic field is increased (the Paschen-Back effect), the polarization of resonance radiation will be very sensitive to changes in the field strength, and the effect will be most marked in the range of field strengths for which the Zeeman splitting of the line becomes of the same order of magnitude as the hyperfine structure itself. Consequently, a measurement of the degree of polarization at various field strengths leads to reliable values of very small hyperfine separations. For the sodium D-lines, the polarization changes from 16 per cent at zero field to 50 per cent at 300 gauss, whence it is calculated that the 3^2P_3-term has a hyperfine separation of 0·000745 cm^{-1}—some twenty times too small to be observed by the usual methods.

In general, the calculation of the polarization to be expected for various I-values is so laborious that it is perhaps better to regard the method as a powerful means of checking known nuclear moments and of extending hyperfine structure analyses.

Good quantitative agreement has been obtained between theory and experiment for sodium, cadmium, caesium and mercury, but in thallium the agreement is only qualitative. In this connection, it should perhaps be pointed out that the hyperfine structures show that the ground states of Tl I are probably very strongly perturbed, so that some discrepancy is not surprising.

Independent methods of determining nuclear moments: (b) The magnetic resolution of atomic beams. A very rapid development has taken place in this field since Breit and Rabi[16] made the suggestion, in 1931, that the Stern-Gerlach experiment might be refined so as to measure nuclear moments. Rabi and his co-workers[17, 18, 19] succeeded in obtaining traces of atomic beams that have traversed a magnetic field of the Stern-Gerlach type, which is, however, not sufficiently strong to destroy the coupling between the nuclear moment and the external electronic system. When the "dispersion" of the instrument is increased by lengthening the path in the inhomogeneous magnetic field, the deflected beam is found to show a $(2I+1)$-fold structure. Thus

the *mechanical* moment may be read off from the *number* of components and the nuclear *magnetic* moment calculated from the *separation* of these components.

A further improvement on the original method is effected by passing the atomic beam through a magnetic "velocity selector" before it is resolved by the Stern-Gerlach field.

In this manner the nuclear spin of sodium was determined for the first time, those of caesium and hydrogen confirmed and the magnetic moments of both hydrogen isotopes were accurately measured. This method thus forms a very welcome addition to the means of fixing nuclear moments.

REFERENCES

§ 1. The determination of wave-length standards

(1) Sears and Barrell. *Philos. Trans.* A, **233**, 143 (1934).
(2) Kösters and Lampe. *Phys. Z.* **35**, 224 (1934).
(3) Sears and Barrell. *Proc. roy. Soc.* A, **139**, 202 (1933).
(4) Pérard. *C.R. Acad. Sci.*, Paris, No. 8, p. 727 (Feb. 1934).
(5) Meggers and Humphreys. *Bur. Stand. J. Res.*, Wash., **13**, 293 (1934).
(6) *J. opt. Soc. Amer.* **11**, 301 (1930).
(7) Jackson, C. V. *Proc. roy. Soc.* A, **143**, 124 (1933).
(8) Humphreys. *Bur. Stand. J. Res.*, Wash., **51** (1930).
(9) Jackson. *Proc. roy. Soc.* A, **138**, 151 (1932).
(10) Dingle. *Mon. Not. R. astr. Soc.* **94**, 866 (1934).

§ 2. Isotope effects in the spectra of diatomic hydrides

(1) Kemble, E. C. and Van Vleck, J. H. *Phys. Rev.* **21**, 653 (1923).
(2) Jenkins, F. A. and McKellar, A. *Phys. Rev.* **42**, 473 (1932).
(3) Jeppesen, C. R. *Phys. Rev.* **45**, 480 (1934); **44**, 165 (1933).
(4) Beutler, H. and Mie, K. *Naturwissenschaften*, **22**, 418 (1934).
(5) Beutler, H. *Z. phys. Chem.* B, **27**, 287 (1934).
(6) Mie, K. *Z. Phys.* **91**, 475 (1934).
(7) Ashley, M. F. *Phys. Rev.* **43**, 770 (1933).
(8) Dieke, G. K. and Blue, R. W. *Phys. Rev.* **47**, 261 (1935).
(9) Lewis, G. N. and Ashley, M. F. *Phys. Rev.* **43**, 837 (1933).
(10) Murphy, G. M. and Johnston, H. *Phys. Rev.* **45**, 500 (1934); **46**, 95 (1934).
(11) Crawford, F. H. and Jorgensen, T. (Junr.). *Phys. Rev.* **47**, 358 (1935); **47**, 932 (1935).
(12) Olsson, E. *Z. Phys.* **93**, 206 (1935).
(13) Koontz, P. G. *Phys. Rev.* **47**, 641 (1935).
(14) Watson, W. W. *Phys. Rev.* **46**, 319 (1934); **47**, 27 (1935).
(15) Guntsch, A. *Z. Phys.* **93**, 534 (1935).
(16) Svensson, E. *Die Bandenspektren des Cadmium-hydrides und deuterides.*
(17) Holst, W. and Hulthén, E. *Z. Phys.* **90**, 712 (1934).
(18) McKellar, A. and Bradley, C. A. *Phys. Rev.* **47**, 787 (1935).
(19) Dieke, G. K. and Blue, R. W. *Phys. Rev.* **45**, 395 (1934).
(20) Johnston, H. L. and Dawson, D. H. *Naturwissenschaften*, **21**, 495 (1933); *Phys. Rev.* **44**, 1031 (1933).
(21) Johnston, H. L. *Phys. Rev.* **45**, 79 (1934).
(22) Shaw, R. W. and Gibbs, R. C. *Phys. Rev.* **45**, 124 (1934).
(23) Hardy, J. D., Barker, E. F. and Dennison, D. M. *Phys. Rev.* **42**, 279 (1932).
(24) Kronig, R. de L. *Physica*, 's Grav., **1**, 617 (1934).

(25) Kronig, R. de L. *Z. Phys.* **50**, 347 (1928).
van Vleck, J. H. *Phys. Rev.* **33**, 467 (1929).
(26) Dunham, J. L. *Phys. Rev.* **41**, 721 (1932).
(27) Dieke, G. K. *Phys. Rev.* **47**, 661 (1935).

§ 3. Astronomical spectroscopy

(1) Grotrian. *Z. Astrophys.* **8**, 124 (1934).
(2) Goudsmit and Ta-You Wu. *Astrophys. J.* **80**, 154 (1934).
(3) Beutler. *Z. Astrophys.* **9**, 387 (1935).
(4) Babcock and Babcock. *Publ. astr. Soc. Pacif.* **46**, 132 (1934).
(5) Moore, Babcock and Kiess. *Astrophys. J.* **80**, 59 (1934).
(6) Moore and Babcock. *Astrophys. J.* **79**, 492 (1934).
(7) King. *Astrophys. J.* **80**, 124 (1934).
(8) Devaux. *C.R. Acad. Sci.*, Paris, **198**, 1595 (1934).
(9) Adel, Slipher and Barker. *Phys. Rev.* **47**, 580 (1935).
(10) Ramanathan and Ramdas. *Proc. Ind. Acad. Sci.* A1, 308 (1934).
(11) Eropkin. *Phil. Mag.* **18**, 838 (1934).
(12) Meyer, Schein and Stoll. *Nature*, Lond., **134**, 535 (1934).
(13) Hunter. *Mon. Not. astr. Soc.* **94**, 594 (1934).
(14) McCrea and Mitra. *Observatory*, **57**, 379 (1934).
(15) Cabannes. *C.R. Acad. Sci.*, Paris, **198**, 2132 (1934).
(16) Garrigue. *C.R. Acad. Sci.*, Paris, **198**, 1398 (1934).
(17) Vegard and Tonsberg. *Z. Phys.* **94**, 413 (1935).
(18) Dufay. *J. Phys. Radium*, **5**, 523 (1934).
(19) Dobrotin, Frank and Čerenkov. *C.R. Acad. Sci. U.S.S.R.* **1**, 110 (1935).
Chvostikov and Lebedev. *C.R. Acad. Sci. U.S.S.R.* **1**, 118 (1935).
(20) Cabannes and Dufay. *C.R. Acad. Sci.*, Paris, **200**, 878 (1935).
(21) —— —— *C.R. Acad. Sci.*, Paris, **200**, 1504 (1935).
(22) Vegard. *Geofys. Publ.* **10**, 4 (1933).
(23) Kaplan. *Phys. Rev.* **45**, 671 (1934).
(24) Williams. *Astrophys. J.* **79**, 280 (1934).
(25) Thorndike. *Lick Obs. Bull.* No. 461, 9 (1934).
(26) Struve. *Astrophys. J.* **79**, 273 (1934).
(27) Beals. *Mon. Not. R. astr. Soc.* **94**, 663 (1934).
(28) Joy. *Publ. astr. Soc. Pacif.* **46**, 51 (1934).
(29) Eddington. *Mon. Not. R. astr. Soc.* **95**, 2 (1934).
(30) Campbell. *Bull. astr. Obs. Harv.* No. 898, 20 (1935).
(31) Dufay and Bloch. *C.R. Acad. Sci.*, Paris, **200**, 217 (1935).
(32) Belorizky. *C.R. Acad. Sci.*, Paris, **200**, 528 (1935).
(33) Stratton and Beer. *Mon. Not. R. astr. Soc.* **95**, 432 (1935).

§ 4. Infra-red band spectra of organic molecules

(1) Julius. *Verh. Akad. Wet.*, Amst., **1**, No. 1 (1892).
(2) Coblentz. "Investigations of infra-red spectra." *Publ. Carneg. Instn* (1905), No. 35.
(3) Schaefer-Matossi. *Das ultrarote Spektrum* (Berlin, 1930): see list of references there.
(4) Lecomte. *Le spectre infrarouge* (Paris, 1928).
(5) Bartholomé. *Z. phys. Chem.* B **23**, 152 (1933).
(6) Levin and Meyer. *J. opt. Soc. Amer.* **16**, 137 (1928).
(7) Ellis. *J. opt. Soc. Amer.* **8**, 1 (1924).
—— *Phys. Rev.* **23**, 48 (1924).
(8) —— *Phys. Rev.* **32**, 906 (1928).
(9) Bartholomé and Teller. *Z. phys. Chem.* B, **19**, 366 (1932).
(10) Lecomte. *C.R. Acad. Sci.*, Paris, **198**, 65 (1934).
(11) Dennison. *Rev. mod. Phys.* **3**, 280 (1931).
(12) Nielsen. *Phys. Rev.* **46**, 117 (1934).

§ 5. THE HYPERFINE STRUCTURE OF SPECTRUM LINES AND NUCLEAR MOMENTS

(1) SCHÜLER, H. and GOLLNOW, H. *Z. Phys.* **93**, 611 (1935).
(2) MORE, K. R. *Phys. Rev.* **46**, 470 (1934).
(3) ANDERSON, O. E. *Phys. Rev.* **46**, 473 (1934).
(4) BALLARD, S. S. *Phys. Rev.* **46**, 806 (1934).
(5) KOPFERMANN, H. and RINDAL, E. *Z. Phys.* **87**, 7 (1934).
(6) TOLANSKY, S. *Proc. roy. Soc.* **144**, 574 (1934).
(7) JONES, E. G. *Proc. roy. Soc.* **144**, 587 (1934).
(8) BROWN, F. W. and BARTLETT, J. H. *Phys. Rev.* **45**, 527 (1934).
(9) JACKSON, D. A. and KUHN, H. *Nature*, Lond., **134**, 25 (1934).
(10) SHENSTONE, A. G. and RUSSELL, H. N. *Phys. Rev.* **39**, 415 (1932).
(11) LANDÉ, A. *Phys. Rev.* **46**, 497 (1934).
(12) SCHÜLER, H. *Z. Phys.* **88**, 323 (1934).
(13) LARRICK, L. *Phys. Rev.* **46**, 581 (1934).
(14) ELLETT, A. and HEYDENBURG, N. P. *Phys. Rev.* **46**, 583 (1934).
(15) HEYDENBURG, N. P. *Phys. Rev.* **46**, 802 (1934).
(16) BREIT, G. and RABI, I. I. *Phys. Rev.* **38**, 2082 (1931).
(17) RABI, I. I. and COHEN, V. W. *Phys. Rev.* **46**, 707 (1934).
(18) COHEN, V. W. *Phys. Rev.* **46**, 713 (1934).
(19) RABI, I. I., KELLOGG, J. M. B. and ZACHARIAS, J. R. *Phys. Rev.* **46**, 157 (1934).

OPTICS

Edited by T. SMITH, M.A., F.R.S.

THIS report may appropriately begin with a word of explanation and apology. In any one year it is only possible to deal with a small selection of topics likely to prove of interest, and the subjects discussed should thus tend to change to some extent from year to year. Arrangements made earlier in the year for contributions on a number of subjects not included in last year's report have all, for one reason or another, proved incapable of fulfilment. It is hoped that articles breaking fresh ground will be available for another year. In their absence this report is confined to subjects which were to some extent considered last year.

§ 1. THE THEORY OF OPTICAL INSTRUMENTS

By T. SMITH, M.A., F.R.S.

The year that is in general receiving special attention in these reports has not been marked by any publications of outstanding importance. When highly important papers do appear they are most frequently of so specialized a nature that they cannot be discussed to advantage in general terms. It will therefore be of wider interest to consider tendencies which have been apparent in recent work.

The additions made in recent years to the more advanced parts of the theory of optical instruments largely relate to ways of partially correcting for aberrations. The interest of partially corrected systems arises from two causes; one is the demonstrable fact that an optical instrument can in general only have its aberrations corrected for a single selected magnification, and the other a belief, or perhaps a fear, that even the theoretically possible state of complete correction for only one magnification may not prove realizable. As the investigations themselves are of a type which cannot be discussed in detail in a general review, only broad principles will be considered here.

When we say that a particular kind of imagery is possible or impossible we mean that it is either consistent or inconsistent with the law of stationary path. This law is implicit in the use of Hamiltonian functions to represent the time of passage of light between certain points or planes. If therefore we can construct a Hamiltonian function of the type corresponding to an optical instrument and find that it implies the absence of aberrations over a particular pair of conjugate surfaces, we say that this kind of correction is possible. Some workers have started on a different basis, but if correct it can be translated into Hamiltonian terms, and the latter is without doubt the simpler starting point. A further difference between investigations arises when the effects of aberrations are considered. Some workers take a geometrical concept such as the cone within which emergent rays lie and the size and shape of its intersection with an image surface as a measure of the defect resulting from the

presence of aberrations. In certain classes of instrument this measure is defensible if the defects are on a somewhat gross scale. As the object however is to derive the conditions which result in minimum defects, it will often be more suitable to take the range over which the phase varies at a mean image point, rather than some quantity based on ray positions, as a measure of the residual effects. The two cases ought to be carefully distinguished because, although the same conditions are reached for the complete absence of aberrations, the measures derived for the seriousness of the defects due to uncorrected aberrations are very different. This difference leads to divergent conclusions on the most effective methods of minimising the effects of those aberrations which cannot be removed.

A distinction must be made between criteria adopted as measures of the residual defects and ways in which the results obtained are expressed. Most work on the design of optical instruments is carried out by tracing rays, trigonometrically or otherwise, through the system, and for this reason it will usually be preferred to express the results in terms of relations between the positions and directions of incident and emergent rays. To take a simple case, we can express in terms of failure to satisfy the Abbe sine law either the variation in magnification near the axis for light traversing different zones of a lens or the differences of phase which will be found at the theoretical image point.

Much theoretical work on possible states of correction has been carried out in recent years, but it is not known how far it has proved applicable to real instruments. When the vast amount of numerical work that has been carried out by optical firms is considered, it seems reasonable to suppose that if precise correction of a plane image of a plane object were really attainable it would have already been found. It appears at least reasonable to suppose that it would have been attained if the method of designing by tracing rays were adequate for the purpose. In fact, notwithstanding the very real advances made in enlarging apertures without reducing the angular field in photographic lenses and some other instruments, the outstanding defects are measurable without difficulty. Herzberger, presumably largely on this ground, has formed the belief that it is impossible with any real instrument to reach perfect correction. (By perfect correction we are to understand correction over a plane in a way consistent with Hamiltonian restriction, and not the impossible standard usually implied when a perfect optical instrument is mentioned.) He admits that he can offer no proof in support of this belief. To the writer it seems that the conclusion is at least premature. Nearly all the systems designed for commercial production are subject to the special restriction that no surface more general than a sphere is to be employed. The evidence can therefore only be interpreted to mean that the chance of its being possible to construct a perfectly corrected lens with only plane or spherical surfaces is extremely small. When surfaces of a more general type are considered, the traditional method of ray tracing offers little attraction. The problem is one which we can expect to receive greater attention in the near future, for our knowledge in this field is surprisingly small. A possible line of procedure leads to conditions in the form of rather complicated differential equations of the first order in several variables which may not all be independent. A useful solution will depend

on the existence of non-trivial general solutions satisfying some additional restrictions.

It may be observed that the reason why Hamilton's Characteristic Function, which is so powerful a weapon in other directions, is not directly helpful in resolving a question of this kind lies in its very virtues. The Characteristic Function is of value precisely because when we employ it we need know nothing whatever about the construction of the instrument we are considering. It only contains those terms which are permissible, and in general we have no need to express the coefficients of its terms as functions of the constructional data of the instrument. If it is in fact impossible to devise a system completely corrected for a given magnification, the coefficients must satisfy some identity or identities which are inconsistent with the values we desire to assign to them. An alternative method of attacking the question is thus to ascertain whether or under what conditions identities connect the values of these coefficients. Our present knowledge goes little beyond the relations which hold when the system is thin.

We have already said that by a perfect optical instrument we must understand one which is corrected for a particular pair of surfaces, and not for three-dimensional space. The incompatibility of the last mentioned state of affairs with the law of light propagation has long been known, and Gullstrand in a celebrated series of papers has argued that we should discard theories which are based upon this impossible ideal and has put forward an alternative. Recently Herzberger has criticized reference to the three-dimensional collinear system as a standard for aberrations. He holds that the standard should be a state which is theoretically attainable. While the present writer entirely agrees with this view, it must be admitted that difficulties do arise from it. These will be very apparent to students who have been taught to regard Gaussian optics as the basis of their work. The reference of all properties to a set of definite planes which have no reference to the special uses for which the instrument is designed has many merits to the teacher who uses this system. The substitution at a later stage of curved surfaces depending on the correction desired is almost sure to prove bewildering. The difficulty will be the less the more the teacher emphasizes the importance of regarding the Gaussian theory as an approximation only definitely valid for light which travels near to the optical axis.

A practical aspect of the inconsistence of corrections for objects at different distances from a lens is not without interest. Some thirty years ago there was a good demand for so-called "universal" photographic objectives, and for "convertible sets", consisting of similar but unequal half-lenses which could be combined variously to yield complete lenses of different focal lengths. These lenses almost certainly owed their origin to belief in the adequacy of the Gaussian concept of imagery. In fact, as that ideal is not attainable, these lenses could not give altogether satisfactory results if their apertures were increased beyond a certain point. It happened that for a considerable time the need to secure larger apertures was not urgent. When the demand for faster lenses became insistent, new types of lens had to be introduced. The same factors which necessitated this change have shown again in recent times but have manifested themselves in a different way. Apertures have

increased to an amazing extent, and to minimize the effects of the conflicting conditions for correct imagery at different distances the focal lengths of the lenses used have been greatly reduced, thus throwing almost all objects to be included in the picture as nearly as may be almost to infinity, where the depth of focus is great. In this way the demand for increased speed has led inevitably to the development of cameras taking smaller pictures. This in turn has made it possible for the film capacity measured in number of exposures to be increased. Further essential conditions of success have been improvements in the film so that great enlargement of the picture could be made, and the fact that the lens aperture being very large—the factor which led to the whole movement—of itself enables finer detail to be clearly recorded on the film if the optical design is of high quality.

§2. THE TEACHING OF ELEMENTARY OPTICS

By T. Smith, M.A., F.R.S.

The publication by the Physical Society of definite recommendations on the conventions to be used in optical work, and particularly in school and university teaching, where diversity of counsels has tended to confusion, will, it is hoped, exercise a lasting influence on theoretical optics in this country and perhaps elsewhere. Since the report appeared the writer has received repeated enquiries on the extent to which schools are being guided by the report. It is clear that many physicists and others interested in scientific education are most anxious that the recommendations contained in the report should be generally adopted, and that one or other of the two sets of subsidiary conventions should also be followed. Unfortunately there is little information available regarding what is happening in the schools. A few letters expressing appreciation of the advantages of the less known Group II convention have been received, and the writer happens to know that in one or two schools it is being used with promising results. As progress in teaching physics forms part of progress in physics itself, it is to be hoped that some material will be forthcoming from the schools on their experiences in this direction. Obviously many who take no part in teaching or in optical investigations are interested.

A reader of one of the older books on optics, such as Robert Smith's *Compleat System of Opticks*, cannot fail to be struck by extensive differences between the older and the newer presentations. The modern emphasis is on object and image, and especially on real images of real objects. The older books gave much attention to concepts which are hardly mentioned in recent books—for instance the idea of apparent distance. This change clearly corresponds to the great importance attained in modern times by instruments which form an image on a chemical or physical receiver. In earlier times the only important instruments were essentially aids to vision, and the images they presented to the observer were virtual. These instruments still form the most important group of optical goods. However easy it is to appreciate the reasons which have led to the change, the neglect of ideas which proved their worth so decisively in the early development of optical theory is to be

regretted. It is certainly the case that useful additions to our knowledge could be made by reintroducing apparent distance as an important concept into our theories. It has just been mentioned in recent theoretical work, and is unquestionably of practical importance. It should not fail to find a place in an elementary school course.

§ 3. VISION

By J. Guild, A.R.C.S., F.Inst.P.

In the retinas of human beings, frogs, and some other creatures is found a bluish red fluid, known as visual purple. The properties of this fluid and its function in the process of vision have been the subject of much speculation and dispute since its discovery by Boll in 1876. In 1879 it was obtained in solution by Kühne, who studied its behaviour when exposed to light. The effect of light is to cause a gradual bleaching of the visual purple, which eventually becomes quite colourless. Kühne observed that at an intermediate stage the colour was yellow. Like later observers, Kühne convinced himself by various tests that this colour was due, not to dilution of the still unbleached purple by the bleached colourless end product, but to the formation of an intermediate substance with distinctive properties of its own. This second substance he termed the visual yellow. On Kühne's hypothesis, the first stage in the decomposition of visual purple is the formation of visual yellow, which is more stable than the original visual purple, and requires light of greater intensity to effect its further decomposition into the colourless product of complete bleaching.

A. König[1], accepting without question both the fact of the formation of visual yellow and Kühne's hypothesis regarding its origin, first determined the spectral absorptions of both the unbleached visual purple and the visual yellow. Most of the work on this subject has, for obvious reasons, been performed with purple extracted from other creatures, usually frogs, but König's results are specially interesting for the reason that the visual purple used in his experiments was from a freshly removed human eye. The limited supply of material contained in a single retina was dissolved in gall, the total amount of solution being only about half a cubic centimetre. With this König contrived to make two separate determinations of the spectrophotometric properties of the unbleached visual purple, of the visual yellow produced in the course of bleaching, and of the end product after complete bleaching.

Various assumptions had to be made and corrections applied before the figures obtained directly in the spectrophotometric measurements could be reduced to yield the absorptive properties of these substances in the thicknesses and concentrations which occur in the retina. The uncertainties introduced in the course of these calculations were probably considerably greater than those due to errors in the spectrophotometry. König succeeded, however, in obtaining curves which, he claimed, represent the spectral absorption of the two substances, visual purple and visual yellow, as they exist in the retina. These curves exhibit two very suggestive coincidences. The spectral absorption curve for visual purple was found to be very

similar to the curve which represents both the brightness sensitivity of normal persons at the very low brightnesses near the threshold of vision and the brightness sensitivity of totally colour blind persons at any brightness level. The spectral absorption curve for visual yellow proved to be somewhat similar to the curve derived by König to show the spectral distribution of blue sensitiveness for normal colour vision.

This blue sensitiveness curve was deduced by applying various corrections to the blue member of König's fundamental sensation curves. To explain this we must digress for a moment. König and Dieterici had studied the colour vision of normal individuals and also of four abnormal observers (dichromats). They succeeded in expressing the results for normal vision by means of three mixture curves having the special property that two of them (the green and the blue or the red and the blue) also represented the colour vision of the abnormal observers. They believed that the trichromatic mixture curves obtained in this way are of fundamental significance in the theory of colour vision. The fact that the vision of all their observers, normal and abnormal alike, could be represented by the combination of some or all of the single set of mixture curves was interpreted as evidence that these three curves should be identified with the spectral sensitivity curves of the three receptor mechanisms which, on König's view of the Young-Helmholtz theory, give rise to the three fundamental sensations of red, green and blue.

The mixture curves, as originally determined, were expressed in terms of stimulus intensities measured before the light entered the eye. The true spectral sensitivity of the retinal receptors has to be expressed in terms of the energy of light of the wave-length considered actually reaching the photosensitive elements, and corrections must therefore be made to the original curves to compensate for any selective absorption which the light may suffer in the eye before these elements are reached. König could not obtain information on the discoloration of the lenses of his six observers, but made corrections for the selective absorption by the yellow spot, and for the spectral energy distribution of the white light employed in his colour matching experiments, and so obtained what he regarded as a fair approximation to the spectral sensitivity curve of the blue-sensitive receptors.

This curve, as we have said, was fairly similar to the absorption curve for visual yellow. König considered that the differences could be attributed to uncertainty in the blue-sensitivity curve, and judged that the evidence was strong enough to establish the conclusion that the two curves, if known accurately for the same eye, would agree very closely. He argued that the twofold similarity of the spectral absorption curve for visual purple to the threshold sensitivity curve, and of the absorption curve for visual yellow to the blue-sensitivity curve, could not be mere coincidence, but indicated that visual purple is the "percipient substance" for light of low intensities below the level at which colour differences appear, and visual yellow that for the perception of blueness. This leaves the nature and origin of percipient substances corresponding to the green and red sensations unexplained, but as we are concerned here with visual purple and its supposed derivative, visual yellow, we shall not stop to consider König's speculations in this region. When due

allowance has been made for the crudity of König's experimental equipment when compared with modern standards, and for the uncertainty involved in the assumptions he had to make before a basis of comparison could be reached, it still remains true that the correspondence in shape and spectral position of his two absorption curves with the two sensitivity curves is sufficiently striking to give considerable plausibility to König's contention that the explanation of the latter is to be sought in the former.

König's view of the function of the visual yellow clearly implies not only that this substance is formed in the retina of the living eye, but also that the proportion in which it is present when the eye has attained any given steady state of adaptation is independent of the events which have led to this state.

The second of these implications follows from the fact that for a given state of adaptation the colour vision is determined only by that state, and does not depend, so far as is known, on whether the eye has reached that state by adaptation from higher or lower brightness levels, or by rapid or slow rates of adaptation. It is necessary to note that there is some evidence against both implications. In discussing the bleaching of visual purple Helmholtz admitted that visual yellow may possibly be formed in the living eye, but regarded it as doubtful; he described an experiment intended to settle the point, but the results were inconclusive. The experiment was based on the view, for which some results of Garten were quoted, that, in the bleaching of retinal extracts, the visual yellow is formed only during rapid bleaching by very strong illumination and not in the course of the slower bleaching which occurs under moderate illumination. On this point the evidence is conflicting, for König's visual yellow was obtained by slow bleaching with green light from a spectroscopic monochromator—a very moderate illumination in no way comparable with the sunlight which Helmholtz believed necessary for the production of visual yellow.

Thus we must admit on the one hand that direct evidence of the formation of visual yellow in the living eye is wanting, and on the other that the evidence against its formation is inconclusive. The first implication of König's theory is not then inconsistent with any established facts. The second implication presents greater difficulty. If the concentration of visual yellow, when the eye is in a given steady state of adaptation, is the same whether this state has been reached from a higher or a lower state, it follows, on the view that this substance is produced from the decomposition of the visual purple, that the concentration of the visual purple also must be independent of the approach to this state; and on either recovery in the dark or adaptation from higher to lower intensities of illumination the colour of the mixture should retrace the course it followed during bleaching. This does not appear to be the case. According to Helmholtz the colours at various stages of the regenerative process are markedly different from those observed during bleaching: the differences described are so great as to leave little doubt of their reality. The regenerative process then almost certainly is not a simple reversal of the bleaching process. This seems at first sight to render untenable the theory that the visual yellow can play an important part in colour vision.

Nevertheless the close similarity between König's absorption curve for this substance and his blue sensitivity curve remains an accepted fact, and while similarities of this nature do not in themselves prove anything, we must not too hurriedly dismiss them as mere chance coincidences. It is possible that a way out of the dilemma will be found as an indirect result of recent work by Hecht and Chase[2]. This work was primarily directed to the resolution of another difficulty—the fact that various observers, including Hecht himself, had failed to obtain evidence of the formation of any distinctive substance corresponding to visual yellow during the bleaching of visual purple in solution. No satisfactory reason for behaviour so completely contradicting that reported by Kühne, König and others was forthcoming; the explanation most favoured attributed the different types of bleaching to effects produced by the particular solvents used or to the methods of purification employed. Hecht and Chase have now thrown new light on this problem by an examination of the bleaching properties of visual purple extracted from frogs at different seasons of the year. The purple was dissolved in bile salts and the method of preparation of the extracts was the same in every case. It was found that extracts taken from frogs in spring or summer showed practically no traces of visual yellow on bleaching, but those taken in the winter showed it in profusion. The frog is a hibernating animal, and hibernation is always accompanied by marked physiological changes. It appears, therefore, that the formation of visual yellow depends on physiological conditions which may vary with the conditions of life of the animal; this circumstance would readily account for the conflict of evidence regarding the reality of the substance. The fact that it may or may not be present in the visual purple of frogs makes it dangerous to draw any conclusions on its presence or absence in human beings, who are not subject to like seasonal physiological changes. Special significance is therefore restored to König's work as establishing the existence of visual yellow in the living human retina, though the view that it is essentially a post mortem product is widely held.

The facts now brought to light about the frog do, however, throw doubt on Kühne's theory that visual yellow is an intermediate decomposition product of visual purple. Hecht and Chase argue that the visual purple itself, and its manner of bleaching, are unlikely to be affected by seasonal physiological changes, and that since the production of visual yellow is not an inevitable result of bleaching visual purple, the former cannot be a decomposition product of the latter, but must rather be derived from some other substance or substances present in the extractions. The visual purple is always found, whatever the condition of the frog, but these other substances, which are assumed to depend on the conditions, may or may not be present.

Despite some slight difference in colour it appears to be generally accepted that the visual purple of man is essentially the same substance as is found in the frog. If this is the case it is also probable that the same holds for the visual yellow, and following Hecht and Chase's argument on the frog, we are led to the view that the visual yellow of the human eye is not produced by the decomposition of visual purple but is attributable to the presence of some other substance or substances.

The physiological conditions for producing these substances must then regularly prevail in man, though in the frog they only hold during hibernation.

If the visual yellow and visual purple be independent substances a way of escape is opened to us from the difficulty following our rejection of the second implication of König's theory of their functions. Their concentrations need no longer be interrelated as they would be if the one substance were a decomposition product of the other: it would be possible for a given concentration of the yellow to coexist with different concentrations of the purple; such differences might account for the different colour sequences observed during bleaching and regeneration. This would not be feasible, even though the substances were independent in origin, if the changes produced by illumination resembled those of a reversible photo-chemical action, for in a state of equilibrium the concentration of each substance would then depend only on the intensity of the illumination, and it would be immaterial whether that illumination were reached from above or below. There is, however, evidence that the changes in visual purple, at least, are not of this reversible type. While the bleaching of visual purple proceeds freely both in solution and in retinas from which the pigment epithelium has been stripped, the regenerative process only takes place in the presence of the pigment layer. Moreover, while bleaching is affected but little by temperature, regeneration is enormously retarded or even stopped altogether in an eye which is allowed to grow cold. The recovery of the purple is not therefore a simple reversal in darkness of the dissociation produced by light. These facts are more in accordance with the view that the visual purple, once bleached, is permanently destroyed, and that the removal of the products of dissociation and the supply of new purple are effected by metabolic processes associated with the pigment epithelium. This theory finds additional support from the fact that the presence of vitamin A in the retina has been established and its influence on the regeneration of visual purple indicated.

If then the regeneration is entirely or largely effected by metabolic processes, the concentration of visual purple corresponding to any given illumination will be determined by the equality of the rates of photic dissociation and of metabolic generation. We may well suppose that the relation between the rates will be considerably affected by the consequences of previous exposure to a strong light on the one hand or of comparative rest on the other. The persistence of such effects for a time would explain, qualitatively at least, differences, dependent on the preceding level of illumination, in the concentration of visual purple for a given intensity of illumination.

On the changes undergone by the visual yellow we have much less information. As will be clear from what has already been said, all we can assert with certainty is that the substance exists and is photosensitive, for like visual purple it loses its colour on prolonged exposure to strong illumination. Of the way in which the colour changes during bleaching or recovery we have no account. If we assume that the equilibrium concentration depends only on the illumination at the moment and is independent of earlier values, we shall not fail to notice that König's identification of visual yellow as the percipient substance for blueness is in harmony with the

observed fact that colour vision in a particular state of steady adaptation is independent of the path by which this state has been reached.

This condition would be satisfied if the dissociation of the visual yellow were strictly a reversible photochemical action, requiring for the regenerative process darkness alone. There is no direct evidence either for or against this hypothesis, but if true it would remove the difficulties from König's theory which the irreversibility of the colour changes in the retinal fluids as a whole at first seemed to impose. The fact that, on this view, the absorption of the visual purple is not uniquely determined by the illumination would not affect colour vision, for the reason that twilight vision, for which the purple is responsible, makes a negligible contribution to the whole sensation at those illuminations for which colour vision is well developed.

Hecht's theory of the independence of visual yellow and visual purple also removes another difficulty from König's theory which the latter attempted to remove, though not with much success. Regarding the yellow as a dissociation product of the purple, König took it for granted that its presence, like that of the purple itself, would be confined to the rods. His theory therefore required that the rod-free area at the fovea should be blue-blind, and he claimed to demonstrate that this was indeed the case. The considerations adduced were far from convincing, and whether or no there exists a minute foveal area comparatively insensitive to blue the fact remains that complete trichromatic colour vision exists for immediate parafoveal regions which are still almost rod free and unaffected to any appreciable extent by the Purkinje effect. If as now seems probable, the visual yellow and visual purple are separate substances, there is no reason to assume that they can only be found in the same place. The yellow may permeate the cones, where its feeble coloration would prevent its being seen.

It is clear that this work of Hecht and Chase has opened up several new roads of enquiry, and some of these may well prove fruitful in extending our knowledge of the processes of vision. It has already invested the old speculations of König, of which very little notice has been taken in recent years on account of the very contradictory statements made by different workers about visual yellow, with fresh interest. Whether König's hypothesis will survive experimental tests cannot be foretold, but his theory is of special interest as the only attempt yet made to explain the spectral characteristics of some of the visual functions from the observed physical properties of retinal constituents. Though it offers no explanation of the perception of green and red by normal trichromats, it is well enough supported, especially after the removal of major difficulties by Hecht's evidence of the independence of visual purple and visual yellow, to add considerable weight to the view that the initial effect of light in vision is to set up photochemical reactions in photo-sensitive substances.

In papers reviewed in the last *Report* Hecht and his collaborators evolved an explanation of the properties of intermittent vision from the theory of photochemical reactions. In a more recent paper Hecht[3] endeavours to explain on a similar basis the law relating the brightness discrimination factor—the so-called Weber or Fechner fraction—with intensity. He assumes that the photochemical changes

suffered by the light sensitive substance are, under certain conditions, reversible. If the initial concentration of the original substance be denoted by a and the concentration of the products of dissociation at the moment t by x, so that the concentration of the original substance at this moment is $a-x$, he supposes that the velocity of the process as a whole under the influence of light of intensity I will be expressed by the relation

$$\frac{dx}{dt} = k_1 I\,(a-x)^m - k_2 x^n,$$

where k_1 and k_2 are constants and m and n represent the orders of the photochemical and the dark reaction respectively. When the eye is completely adapted to the intensity I, the rate of decomposition of the photosensitive substance is equal to the rate of its formation by reunion of the photoproducts, and dx/dt is zero. If the intensity is then increased by ΔI, the value of dx/dt, while the eye is still in the state of adaptation corresponding to the initial intensity I, is proportional to $\Delta I\,(a-x)^m$.

In applying these relations to the problem of intensity discrimination two hypotheses are made: first, that the recognition of a change of brightness is determined by the fact that dx/dt is not zero, so that the conditions prevailing before the eye has time to adapt itself to the new brightness are involved, and secondly, that the least value of dx/dt which will cause a change to be recognized is independent of the intensity. On these hypotheses the form of the law connecting I and the least perceptible value of $\Delta I/I$ for various types of reaction can be deduced, and by assuming suitable values of m and n Hecht succeeds in computing curves which fit what is known on brightness discrimination for *Drosophila*, the bee, and for the human eye. The range of brightness covered includes the region of "twilight vision" for which the rods are responsible—that is the conditions under which König supposed the photo-sensitive substance to be the visual purple. The fact that Hecht's treatment assumes a reversible photochemical action though there are good reasons for believing that the dissociation of visual purple is not reversible, does not present serious difficulty. Any method of generation of the photo-sensitive substance which makes the rate of generation depend on the deficiency of the substance compared with its normal concentration in darkness will result in equations generally similar to those for a reversible action. A metabolic generative process must depend primarily on this deficiency, though events in the immediate past may raise or lower the rate. These variations produce irregular departures rather than affect the general form of the law. Our knowledge of brightness discrimination for rod vision is neither extensive nor consistent enough to justify the belief that the same law is always exactly followed.

More serious difficulties arise from Hecht's hypothesis that the recognition of a change of brightness depends on the value of dx/dt which prevails before the eye has had time to become adapted to the new conditions. One of these, the fact that a brightness difference between the juxtaposed fields remains perceptible for an indefinite time, is discussed in the paper, and an explanation based on involuntary wanderings of the point of fixation is offered. Whether these wanderings can fill so important a role has still to be determined. The question appears to be one which

physiologists should be able to settle experimentally. There are other difficulties in the way of accepting this hypothesis which we need not discuss here. They will doubtless be urged in the discussion which this very suggestive paper is likely to provoke. Whether the difficulties are removed or not, the work of Hecht and his school has now made it clear that for an explanation of the phenomena associated with the first stage in vision—the process of robbing the ether waves of their energy prior to its utilization by the neural system for the activities which culminate in the perception of light and colour—we must look to the theory of photo-chemical reactions.

§ 4. COLORIMETRY

By J. GUILD, A.R.C.S., F.Inst.P.

In the short interval between the preparation of the last *Report* and the end of 1934, there have been few papers bearing on colorimetry, and fewer still breaking fresh ground either in technique or instrumental design.

Some new results are contained in a paper by G. Haase[1], who has investigated the sensitivity of the eye to differences in hue-wave-length at various brightness levels and colorimetric saturations. The hue discrimination threshold throughout a spectrum of constant brightness was determined for a series of brightness levels varying from approximately $1{\cdot}4 \times 10^{-10}$ candle per sq. cm. to $1{\cdot}4 \times 10^{-5}$ candle per sq. cm. The sensitivity was found to be practically independent of the brightness level except near the upper and lower limits of this range. The general features of the results are fairly similar to those obtained by previous workers. As regards the secondary maximum of sensitivity in the neighbourhood of 640mμ, the reality of which has been disputed by some writers,* the investigation is inconclusive, as some of the curves exhibit a maximum here while others do not. There is clearly room for further work in this region of the spectrum.

The influence of desaturation with white on hue discrimination was also investigated. No diminution in hue sensitivity was detected until the proportion of white exceeded 50 per cent. Towards the blue end of the spectrum (480–400mμ) there was no observable decrease even when the white amounted to 75 per cent of the whole. These results are in general agreement with those previously obtained by E. P. T. Tyndall. It appears therefore to be fairly well established that the number of liminal steps along any equi-saturation locus on the colour chart is (at least nearly) constant for colorimetric saturations from about 50 to 100 per cent. Investigations of this kind, together with those on the discrimination of differences in saturation, are gradually building up a body of knowledge from which it will be possible to predict the discrimination limen in any direction at any point of the colour triangle. The position has greatly improved in the last few years, for not very long ago systematic experiments on the permissible magnitude of colorimetric errors had only been carried out along the spectral locus.

* For example Wright and Pitt, as noted in the last *Report*, could find no trace of this maximum.

D. L. MacAdam[2] has investigated the factors which determine the degree of "whiteness" of white objects. That is to say, if we are given two objects of high but different reflecting factors and very pale but different colours, he has considered which of the two we shall regard as the "whiter" and why we reach our conclusion. The answer depends partly on the relation between discrimination factors for colour and brightness, and partly on psychological factors. MacAdam concludes that if the colorimetric saturations are less than about 3 per cent, the estimate of "whiteness" is determined by the brightness factor; but, if the saturation exceeds about 5 per cent, the saturation itself and not the brightness factor determines the answer. The relation between the results of objective tests and the partially subjective criterion by which the excellence of white goods is ultimately appraised is of some practical significance in laundering, bleaching and similar operations.

Though no new instruments for direct colour measurement have been described within the period under review, H. P. J. Verbeek[3] has described an improved form of his trichromatic colorimeter mentioned in the last *Report*. The new design employs an optical system which ensures that the exit pupil is uniformly filled with light of each of the three instrumental primaries, as well as with light from the test object. Parallactic effects due to the chromatic aberration of the eye, which had been found troublesome in the original instrument, have thereby been overcome.

New equipment for the measurement of reflection or transmission factors under controlled conditions of illumination has been described by H. J. McNicholas[4]. The equipment includes provision for illuminating the specimen by light of controlled spectral composition, and this may be either diffuse or unidirectional. With either type of illumination the light reflected or transmitted by the sample in any arbitrarily chosen direction may be measured by visual photometric or spectrophotometric methods.

R. D. Nutting[5] has described a preliminary investigation on the utility of the Hardy recording photoelectric spectrophotometer for the discrimination of small colour differences in dyed textiles. Specimens of wool treated with varying proportions of red and yellow dyes were examined and the results interpreted in terms of trichromatic coordinates for the standard observer. The author concludes that the results establish the suitability of the Hardy spectrophotometric method for the detection of small colour differences.

Special systems of colour measurement, some of which were in widespread use before the introduction of the C.I.E. system, naturally continue to be used for some purposes for which their utility has been well established. Further information concerning two of these systems has been published. I. H. Godlove and A. E. O. Munsell[6] have examined the precision which may be obtained by using the Munsell colour charts for describing the colour of a pigmented sample. Thirty-two samples were examined spectrophotometrically and their colorimetric constants computed. The results are recorded in terms of the steps of the Munsell scale and compared with those obtained by matching the samples directly against the Munsell standards. The uncertainty of the experimental determination was found to be

about one Munsell step in hue and half a Munsell step in chroma (equivalent to colorimetric saturation). The agreement in hue was best for blues and worst for yellows and yellow-reds, but these positions were reversed for chroma. The computed results are unfortunately based on the properties of the old O.S.A. standard observer and are not, therefore, as useful as they would otherwise have been in establishing a connection between the Munsell standards and the C.I.E. colorimetric system.

Further work on the red series of the Lovibond Tintometer glasses is reported from the Bureau of Standards. These glasses are used in large numbers in the United States for grading vegetable oils by their colour, and inconsistencies in the grading of the glasses may easily cause discrepancies which are commercially serious. K. S. Gibson and Geraldine W. Haupt[7], have restandardized the series of red Lovibond glasses held at the Bureau of Standards. New numerals have been derived, approximately equal to the original Lovibond numerals, but free from certain inconsistencies which occurred in them. These numerals have been derived from colorimetric computations based on spectrophotometric measurements of the glasses in combination with number 35 yellow, and are on a consistent additive scale. This scale is now used for the grading of all glasses tested at the Bureau of Standards.

A new application of colour matching theory to optical pyrometry is described in the paper by G. Haase[1] to which reference has already been made. In a colour pyrometer which he has designed light from the radiator whose temperature is to be measured is split into four pencils; after passing through suitable colour filters these are recombined in two pairs in a bi-partite matching field. Red and green filters are used for one pair of pencils, orange and blue for the other. The intensities are controlled by absorbing wedges, and can be adjusted so that the two fields match in colour and brightness. From the relative intensities of the pencils when matches are attained the temperature of the radiator, that is to say, its colour temperature if it is not a full radiator, can be deduced.

An account has appeared of work on the colour-temperature scale now established at the National Bureau of Standards. A gas-filled lamp at a colour temperature of 2848° K. is one of the fundamental standards of the C.I.E. system; it is the radiator which, either alone or with specified filters, provides illumination of standard spectral composition for colorimetric measurements. The accuracy with which the colour temperature scale can be realized in the neighbourhood of 2800 to 2900° K. is therefore an important factor in determining the accuracy with which colorimetric measurements can be made in terms of the C.I.E. system. It is the only factor for which the technical colorimetrist has to place reliance on other workers. While the specification of the standard illuminants is precise, and the filters involved in two of them can be made up with high accuracy, the rating of lamps on the colour-temperature scale is naturally a function of national standardizing laboratories. The basis of the colour-temperature scale maintained at the National Physical Laboratory has been described in the Annual Reports of the

Laboratory.* The scale employed until recently at the National Bureau of Standards was based principally on a number of lamps rated at the Nela Research Laboratories. A new scale has recently been established at the Bureau and has been described by H. T. Wensel, D. B. Judd and W. F. Roeser[8]. It is based on direct colour matches between a series of gas-filled lamps and black bodies immersed in freezing platinum, rhodium and iridium. The temperatures of these fixed points are respectively 2046, 2239 and 2727° K. Interpolation between these temperatures is effected by utilizing a simple relation between the colour temperature of the lamp as a whole and the brightness temperature of one of its coils. The relation between colour temperature and voltage for any lamp is represented accurately by the formula $T_c = a + bV^{\frac{1}{2}}$. Extrapolation to temperatures higher than 2727° K. may be effected by the formula, by reference to brightness temperatures of a selected coil, or by the use of suitable liquid filters. The new scale gives higher values for the colour temperature of a lamp at a given voltage than the old scale, the difference being 20° at 2046° K., and 14° at both 2239 and 2727° K.

A preliminary comparison has been carried out[9] of this scale with that of the National Physical Laboratory. The comparison at the lower end of the temperature range was carried out on three lamps. For colour temperatures between about 2000 and 2250° K. the N.P.L. values are lower than those obtained by the N.B.S. by from 5 to 7°. At higher colour temperatures the comparison was based on one lamp only, and should therefore be regarded as less certain. The results obtained indicate that the N.P.L. assigns values about 12° lower than the N.B.S. in the neighbourhood of 2727° K., and about 15° lower in the neighbourhood of 2850° K., the most important colour temperature for colorimetric standardization.

As this is the first intercomparison of the scales established in two different countries, the result, from a practical point of view, must be regarded as highly satisfactory, particularly as a discrepancy of this order has no appreciable effect on the accuracy of technical colorimetry. Nevertheless, the discrepancies are considerably greater than those which now exists in measurements of high temperatures by approved methods of optical pyrometry. As the uncertainty in the colour temperature scale should correspond to the accuracy with which the international temperature scale is realized, there is clearly room for further work on the improvement and correlation of colour temperature scales.

REFERENCES

§ 3. Vision

(1) Konig, A. *S.B. preuss. Akad. Wiss.* 577 (1894).
(2) Hecht, Selig and Chase, Aurin M. "Anomalies in the absorption spectrum of visual purple." *Proc. Nat. Acad. Sci.*, Wash., **20**, 238–43 (1934).
(3) Hecht, Selig. "A theoretical basis for intensity discrimination in vision." *Proc. Nat. Acad. Sci.*, Wash., **20**, 644–55 (1934).

* The colour-temperature scale of the N.P.L. is based on vacuum lamp standards of low colour temperature which have been rated by direct comparison with a Planckian radiator at a series of known temperatures from about 1800 to 2400° K. The step-up to higher colour temperatures characteristic of gas-filled lamps is effected by means of a specially designed colour filter.

§ 4. COLORIMETRY

(1) HAASE, G. "Hue sensitivity of the eye at different brightnesses and saturations. Construction of a sensitive colour pyrometer." *Ann. Phys.*, Lpz., **20**, 1, 75–105 (1934).
(2) MACADAM, D. L. "Specification of whiteness." *J. opt. Soc. Amer.* **24**, 188–91 (1934).
(3) VERBEEK, H. P. J. "Trichromatic colorimeter." *Physica*, 's Grav., **1**, 1082–4 (1934), (in English).
(4) MCNICHOLAS, H. J. "Diffuse reflection and transmission." *Bur. Stand. J. Res.*, Wash., **13**, 211–36 (1934).
(5) NUTTING, R. D. "Detection of small colour differences in dyed textiles." *J. opt. Soc. Amer.* **24**, 135–8 (1934).
(6) GODLOVE, I. H. and MUNSELL, A. E. O. "Colorimetry with a colour atlas." *J. opt. Soc. Amer.* **24**, 267–71 (1934).
(7) GIBSON, K. S. and HAUPT, GERALDINE W. "Standardization of Lovibond red glasses in combination with Lovibond 35 yellow." *Bur. Stand. J. Res.*, Wash., **13**, 433–51 (1934).
(8) WENSEL, H. T., JUDD, D. B. and ROESER, W. F. "Establishment of a scale of colour temperature." *Bur. Stand. J. Res.*, Wash., **12**, 527 (1934).
(9) *Rep. nat. phys. Lab.*, Lond., pp. 53, 81 (1934).